



(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)



1-0051689

(51)^{2006.01} H04M 9/08; G10L 15/22

(13) B

-
- (21) 1-2022-00931 (22) 29/07/2020
(86) PCT/US2020/043958 29/07/2020 (87) WO2021/021857 A1 04/02/2021
(30) P201930702 30/07/2019 ES; 62/880,113 30/07/2019 US; 62/880,122 30/07/2019 US;
19212391.7 29/11/2019 EP; 62/950,004 18/12/2019 US; 62/971,421 07/02/2020 US;
62/705,410 25/06/2020 US; 62/705,897 21/07/2020 US
(45) 25/09/2025 450 (43) 25/04/2022 409A
(73) 1. DOLBY LABORATORIES LICENSING CORPORATION (US)
1275 Market Street San Francisco, California 94103 (US)
2. DOLBY INTERNATIONAL AB (NL)
Apollo Building, 3E Herikerbergweg 1-35 1101 CN Amsterdam Zuidoost (NL)
(72) DICKINS, Glenn N. (AU); HINES, Christopher Graham (AU); GUNAWAN, David
(AU); CARTWRIGHT, Richard J. (AU); SEEFELDT, Alan J. (US); ARTEAGA,
Daniel (ES); THOMAS, Mark R. P. (GB); LANDO, Joshua B. (US).
(74) Công ty TNHH Quốc tế D & N (D&N INTERNATIONAL CO.,LTD.)
-

(54) PHƯƠNG PHÁP, THIẾT BỊ VÀ HỆ THỐNG QUẢN LÝ PHIÊN ÂM THANH, VÀ
PHƯƠNG TIỆN BẮT BIẾN ĐỌC ĐƯỢC BẰNG MÁY TÍNH

(21) 1-2022-00931

(57) Sáng chế đề cập đến phương pháp xử lý âm thanh có thể bao gồm bước nhận tín hiệu đầu ra từ mỗi micrô trong số nhiều micrô trong môi trường âm thanh, tín hiệu đầu ra tương ứng với tiếng nói hiện tại của người và xác định, dựa trên tín hiệu đầu ra, một hoặc nhiều khía cạnh của thông tin ngữ cảnh liên quan đến người đó, bao gồm độ gần hiện tại được ước lượng của người đó với một hoặc nhiều vị trí micrô. Phương pháp này có thể bao gồm bước lựa chọn hai hoặc nhiều thiết bị âm thanh được trang bị loa dựa, ít nhất một phần, vào một hoặc nhiều khía cạnh của thông tin ngữ cảnh, xác định một hoặc nhiều kiểu thay đổi xử lý âm thanh để áp dụng cho dữ liệu âm thanh đang được kết xuất cho tín hiệu cấp cho loa cho các thiết bị âm thanh và khiến cho một hoặc nhiều kiểu thay đổi xử lý âm thanh sẽ được áp dụng. Theo một số ví dụ, các thay đổi xử lý âm thanh có tác dụng tăng tỷ lệ lời nói trên tiếng vọng ở một hoặc nhiều micrô. Sáng chế đề cập đến phương pháp, thiết bị và hệ thống quản lý phiên âm thanh, và phương tiện bất biến đọc được bằng máy tính.

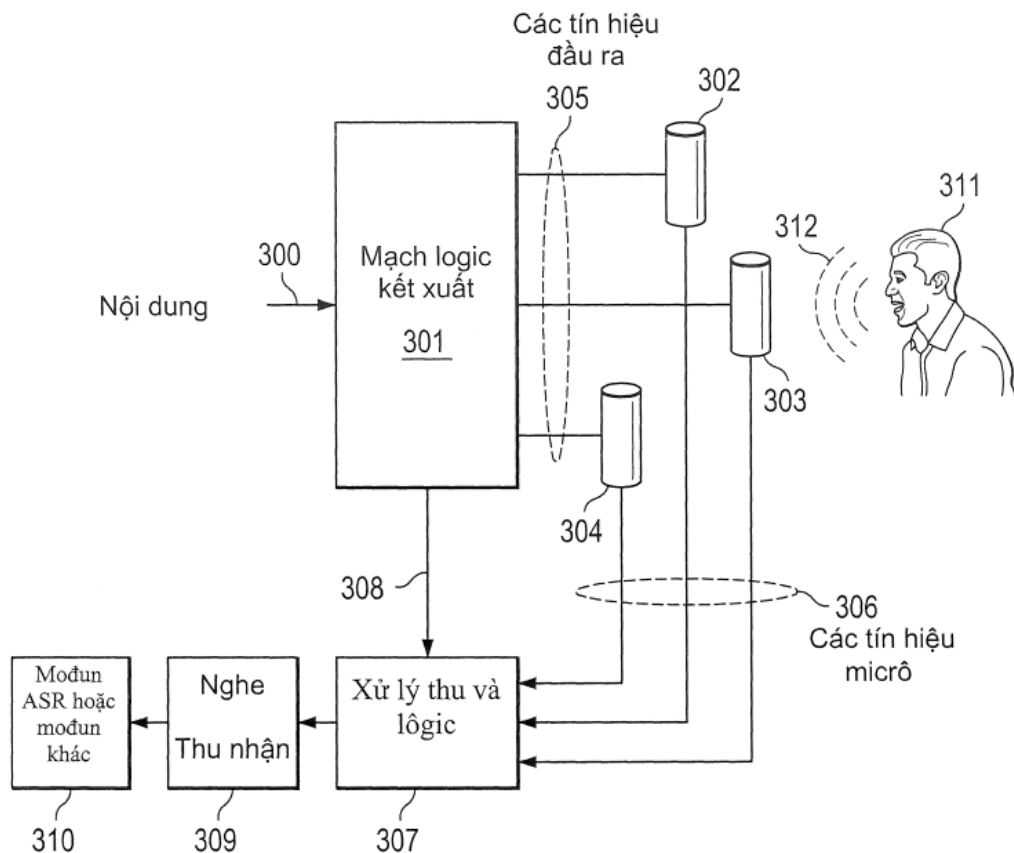


Fig.3A

Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến các hệ thống và phương pháp phối hợp (hòa âm) và ứng dụng các thiết bị âm thanh (ví dụ, thiết bị âm thanh thông minh) và điều khiển việc kết xuất âm thanh bằng các thiết bị âm thanh.

Tình trạng kỹ thuật của sáng chế

Các thiết bị âm thanh, bao gồm nhưng không giới hạn ở các thiết bị âm thanh thông minh, đã được sử dụng rộng rãi và đang trở thành đặc điểm chung của nhiều gia đình. Mặc dù các hệ thống và phương pháp điều khiển thiết bị âm thanh hiện nay có lợi ích, nhưng mong muốn có các hệ thống và phương pháp được cải tiến.

Kí hiệu và thuật ngữ

Trong toàn bộ bản mô tả này, bao gồm cả trong phần yêu cầu bảo hộ, “loa” và “loa phóng thanh” được dùng đồng nghĩa để biểu thị bộ chuyển đổi phát ra âm thanh bất kỳ (hoặc tập hợp bộ chuyển đổi) được điều khiển bằng một nguồn cấp loa. Một bộ tai nghe điển hình bao gồm hai loa. Loa có thể được áp dụng để bao gồm nhiều bộ chuyển đổi (ví dụ, loa trầm và loa cao tần), mà có thể được điều khiển bằng một nguồn cấp loa chung hoặc nhiều nguồn cấp loa. Trong một số ví dụ, (các) nguồn cấp loa có thể trải qua các quy trình xử lý khác nhau trong các nhánh mạch khác nhau được kết hợp với các bộ chuyển đổi khác nhau.

Trong toàn bộ bản mô tả này, bao gồm cả trong phần yêu cầu bảo hộ, thuật ngữ thực hiện hoạt động “trên” tín hiệu hoặc dữ liệu (ví dụ, lọc, định tỷ lệ, chuyển đổi, hoặc áp dụng độ khuếch đại cho, tín hiệu hoặc dữ liệu) được dùng theo nghĩa rộng để biểu thị việc thực hiện trực tiếp hoạt động trên tín hiệu hoặc dữ liệu, hoặc trên phiên bản đã xử lý của tín hiệu hoặc dữ liệu (ví dụ, trên phiên bản của tín hiệu mà đã trải qua quy trình lọc sơ bộ hoặc tiền xử lý trước khi thực hiện hoạt động trên đó).

Trong toàn bộ bản mô tả này, bao gồm cả trong phần yêu cầu bảo hộ, thuật ngữ “hệ thống” được dùng theo nghĩa rộng để biểu thị thiết bị, hệ thống, hoặc hệ thống phụ. Ví dụ, hệ thống phụ thực hiện bộ giải mã có thể được gọi là hệ thống giải mã, và hệ thống bao gồm hệ thống phụ như vậy (ví dụ, hệ thống tạo ra X tín hiệu đầu ra để đáp lại

nhiều đầu vào, trong đó hệ thống phụ tạo ra M đầu vào và X - M đầu vào khác được nhận từ nguồn bên ngoài) cũng có thể được gọi là hệ thống giải mã.

Trong toàn bộ bản mô tả này, bao gồm cả trong phần yêu cầu bảo hộ, thuật ngữ “bộ xử lý” được dùng theo nghĩa rộng để chỉ hệ thống hoặc thiết bị có thể lập trình được hoặc theo cách khác tạo cấu hình được (ví dụ, với phần mềm hoặc firmware) để thực hiện các hoạt động trên dữ liệu (ví dụ, âm thanh, hoặc video hoặc dữ liệu hình ảnh khác). Các ví dụ của bộ xử lý bao gồm mảng cổng lập trình được theo trường (hoặc mạch tích hợp tạo cấu hình được khác hoặc bộ chip), bộ xử lý tín hiệu số được lập trình và/hoặc theo cách khác được tạo cấu hình để thực hiện xử lý đường ống trên âm thanh hoặc dữ liệu âm thanh khác, bộ xử lý hoặc máy tính đa năng có thể lập trình được, và chip hoặc bộ chip của bộ vi xử lý lập trình được.

Trong toàn bộ bản mô tả này, bao gồm cả trong phần yêu cầu bảo hộ, thuật ngữ “ghép nối” hoặc “được ghép nối” được dùng theo nghĩa rộng nghĩa là kết nối trực tiếp hoặc gián tiếp. Do đó, nếu thiết bị thứ nhất ghép nối với thiết bị thứ hai, thì kết nối này có thể là kết nối trực tiếp, hoặc kết nối gián tiếp qua các thiết bị và kết nối khác.

Như được dùng ở đây, “thiết bị thông minh” là thiết bị điện tử, thường được tạo cấu hình để truyền thông với một hoặc nhiều thiết bị (hoặc mạng) khác qua các giao thức không dây khác nhau như Bluetooth, Zigbee, truyền thông trường gần, Wi-Fi, truyền thông bằng ánh sáng (light fidelity - Li-Fi), 3G, 4G, 5G, v.v., mà có thể hoạt động ở một mức độ nào đó có tính tương tác và/hoặc độc lập. Một số loại thiết bị thông minh đáng chú ý là điện thoại thông minh, ô tô thông minh, bộ điều chỉnh nhiệt thông minh, chuông cửa thông minh, khóa thông minh, tủ lạnh thông minh, phablet và máy tính bảng, đồng hồ thông minh, dây đeo thông minh, chuỗi khóa thông minh và các thiết bị âm thanh thông minh. Thuật ngữ “thiết bị thông minh” cũng có thể đề cập đến thiết bị có một số đặc tính của máy tính phổ biến, như trí tuệ nhân tạo.

Ở đây, chúng tôi sử dụng cụm từ “thiết bị âm thanh thông minh” để biểu thị thiết bị thông minh là thiết bị âm thanh chuyên dùng hoặc thiết bị âm thanh đa năng (ví dụ, thiết bị âm thanh áp dụng ít nhất một số khía cạnh của chức năng trợ lý ảo). Thiết bị âm thanh chuyên dùng là thiết bị (ví dụ, ti vi (television - TV) hoặc điện thoại di động) bao gồm hoặc gắn với ít nhất một micro (và tùy ý cũng bao gồm hoặc gắn với ít nhất một loa và/hoặc ít nhất một camera), và được thiết kế hầu như hoặc chủ yếu để đạt được một

mục đích duy nhất. Ví dụ, mặc dù TV thường có thể phát (và được coi là có khả năng phát) âm thanh từ nội dung chương trình, trong hầu hết các trường hợp, TV hiện đại chạy hệ điều hành nào đó mà các ứng dụng chạy cục bộ trên đó, bao gồm cả ứng dụng xem truyền hình. Tương tự, đầu vào và đầu ra âm thanh trong điện thoại di động có thể làm được nhiều việc, nhưng những thứ này được phục vụ bằng các ứng dụng chạy trên điện thoại. Theo hướng này, thiết bị âm thanh chuyên dùng có (các) loa và (các) micrô thường được tạo cấu hình để chạy ứng dụng và/hoặc dịch vụ cục bộ để sử dụng trực tiếp (các) loa và (các) micrô. Một số thiết bị âm thanh chuyên dùng có thể được tạo cấu hình để nhóm lại với nhau để phát âm thanh qua vùng hoặc khu vực được tạo cấu hình bởi người dùng.

Một loại thiết bị âm thanh đa năng phổ biến là thiết bị âm thanh áp dụng ít nhất một số khía cạnh của chức năng trợ lý ảo, mặc dù các khía cạnh khác của chức năng trợ lý ảo có thể được áp dụng bằng một hoặc nhiều thiết bị khác, như một hoặc nhiều máy chủ mà thiết bị âm thanh đa năng được tạo cấu hình để truyền thông.

Ở đây, thiết bị âm thanh đa năng như vậy có thể được gọi là “trợ lý ảo”. Trợ lý ảo là một thiết bị (ví dụ, loa thông minh hoặc thiết bị tích hợp trợ lý giọng nói) bao gồm hoặc kết hợp với ít nhất một micrô (và tùy ý cũng có thể bao gồm hoặc kết hợp với ít nhất một loa và/hoặc ít nhất một camera). Theo một số ví dụ, trợ lý ảo có thể có khả năng sử dụng nhiều thiết bị (khác trợ lý ảo) cho các ứng dụng theo nghĩa được đám mây được hỗ trợ hoặc không được áp dụng hoàn toàn trong hoặc trên chính trợ lý ảo. Nói cách khác, ít nhất một số khía cạnh của chức năng trợ lý ảo, ví dụ, chức năng nhận dạng lời nói, có thể được áp dụng (ít nhất một phần) bởi một hoặc nhiều máy chủ hoặc thiết bị khác mà trợ lý ảo có thể truyền thông qua mạng, như Internet. Các trợ lý ảo đôi khi có thể hoạt động cùng nhau, ví dụ, theo cách rời rạc và được xác định có điều kiện. Ví dụ, hai hoặc nhiều trợ lý ảo có thể làm việc cùng nhau theo nghĩa, một trong số chúng, ví dụ, trợ lý ảo tin chắc rằng đã nghe thấy từ đánh thức, đáp lại từ đánh thức đó. Các trợ lý ảo đã kết nối có thể, theo một số phương án thực hiện, có thể tạo ra một loại chòm sao, mà có thể được quản lý bằng một ứng dụng chính có thể là (hoặc áp dụng) trợ lý ảo.

Ở đây, "từ đánh thức" được dùng theo nghĩa rộng để biểu thị âm thanh bất kỳ (ví dụ, một từ do con người nói ra, hoặc âm thanh khác nào đó), trong đó thiết bị âm thanh

thông minh được tạo cấu hình để thức để đáp lại việc phát hiện ("nghe") âm thanh (bằng cách sử dụng ít nhất một micrô bao gồm trong hoặc kết hợp với thiết bị âm thanh thông minh, hoặc ít nhất một micrô khác). Trong ngữ cảnh này, "thức" biểu thị rằng thiết bị đi vào trạng thái thiết bị chờ (nói cách khác, đang nghe) lệnh âm thanh. Trong một số trường hợp, những từ có thể được gọi ở đây là "từ đánh thức" có thể bao gồm nhiều hơn một từ, ví dụ, một cụm từ.

Ở đây, cụm từ "bộ phát hiện từ đánh thức" biểu thị thiết bị được tạo cấu hình (hoặc phần mềm bao gồm các lệnh để tạo cấu hình thiết bị) để tìm kiếm một cách liên tục sự đồng chỉnh giữa các đặc trưng âm thanh thời gian thực (ví dụ, lời nói) và mô hình được huấn luyện. Thông thường, một sự kiện từ đánh thức được kích hoạt bất cứ lúc nào bộ phát hiện từ đánh thức xác định rằng xác suất một từ đánh thức đã được phát hiện vượt quá ngưỡng xác định trước. Ví dụ, ngưỡng có thể là ngưỡng xác định trước được điều chỉnh để tạo ra sự thỏa hiệp hợp lý giữa tỷ lệ chấp nhận sai và từ chối sai. Sau sự kiện từ đánh thức, thiết bị có thể đi vào trạng thái (có thể được gọi là trạng thái "thức tỉnh" hoặc trạng thái "chăm chú") trong đó thiết bị lắng nghe lệnh và chuyển tiếp lệnh đã nhận cho bộ nhận dạng lớn hơn, có cường độ tính toán lớn hơn.

Ở đây, cụm từ "vị trí micrô" biểu thị vị trí của một hoặc nhiều micrô. Theo một số ví dụ, vị trí của một micrô có thể tương ứng với mảng micrô nằm trong một thiết bị âm thanh. Ví dụ, vị trí micrô có thể là vị trí duy nhất tương ứng với toàn bộ thiết bị âm thanh bao gồm một hoặc nhiều micrô. Theo một số ví dụ như vậy, vị trí micrô có thể là một vị trí tương ứng với trung tâm của mảng micrô của thiết bị âm thanh. Tuy nhiên, trong một số trường hợp, vị trí micrô có thể là vị trí của một micrô. Theo một số ví dụ như vậy, thiết bị âm thanh có thể chỉ có một micrô.

Bản chất kỹ thuật của sáng chế

Một số phương án của sáng chế đề xuất phương pháp quản lý trải nghiệm người nghe hoặc "người dùng" để cải thiện tiêu chí chính để song công toàn phần thành công tại một hoặc nhiều thiết bị âm thanh. Tiêu chí này được gọi là tỷ lệ Tín hiệu trên tiếng vọng (Signal to Echo ratio - SER), còn được gọi là Tỷ lệ lời nói trên tiếng vọng, mà có thể được định nghĩa là tỷ lệ giữa tín hiệu giọng nói (hoặc tín hiệu mong muốn khác) thu được từ môi trường (ví dụ, phòng) qua một hoặc nhiều micrô, và tiếng vọng phát ra từ thiết bị âm thanh bao gồm một hoặc nhiều micrô từ nội dung chương trình đầu ra, nội

dung tương tác, v.v. Dự tính rằng nhiều thiết bị âm thanh của môi trường âm thanh có thể có cả loa và micrô tích hợp trong khi đáp ứng các chức năng khác. Tuy nhiên, các thiết bị âm thanh khác của môi trường âm thanh có thể có một hoặc nhiều loa nhưng không có (các) micrô, hoặc một hoặc nhiều micrô nhưng không có (các) loa. Một số phương án cố ý tránh sử dụng (hoặc chủ yếu không sử dụng) (các) loa gần nhất đối với người dùng trong các trường hợp hoặc tình huống sử dụng nhất định. Ngoài ra, hoặc theo cách khác, một số phương án có thể khiến cho một hoặc nhiều kiểu thay đổi xử lý âm thanh khác đối với dữ liệu âm thanh được kết xuất bằng một hoặc nhiều loa của môi trường âm thanh để tăng SER ở một hoặc nhiều micrô của môi trường.

Một số phương án được tạo cấu hình để triển khai hệ thống bao gồm các thiết bị âm thanh được phối hợp (phối nhạc), mà trong một số phương án có thể bao gồm các thiết bị âm thanh thông minh. Theo một số phương án thực hiện như vậy, hai hoặc nhiều thiết bị âm thanh thông minh là (hoặc được tạo cấu hình để triển khai) bộ phát hiện từ đánh thức. Do đó, nhiều micrô (ví dụ, micrô không đồng bộ) có sẵn trong các ví dụ như vậy. Trong một số trường hợp, mỗi micrô có thể được bao gồm trong, hoặc được tạo cấu hình để truyền thông với, ít nhất một trong các thiết bị âm thanh thông minh. Ví dụ, ít nhất một số micrô có thể là micrô rời (ví dụ, trong các thiết bị gia dụng) không được bao gồm trong bất kỳ thiết bị âm thanh thông minh nào, nhưng được tạo cấu hình để truyền thông với (sao cho đầu ra của chúng là có thể thu được bởi) ít nhất một trong số các thiết bị âm thanh thông minh. Trong một số phương án, mỗi bộ phát hiện từ đánh thức (hoặc mỗi thiết bị âm thanh thông minh bao gồm bộ phát hiện từ đánh thức), hoặc hệ thống phụ khác (ví dụ, bộ phân loại) của hệ thống, được tạo cấu hình để ước lượng vùng của người bằng cách áp dụng bộ phân loại được điều khiển bằng nhiều đặc tính âm thanh thu được từ ít nhất một số micrô (ví dụ, micrô không đồng bộ). Theo một số phương án thực hiện, mục tiêu có thể không phải là ước lượng vị trí chính xác của người đó mà thay vào đó là tạo ra ước lượng mạnh mẽ về một vùng riêng mà bao gồm vị trí hiện tại của người đó.

Theo một số phương án thực hiện, người (ở đây cũng có thể được gọi là “người dùng”), thiết bị âm thanh thông minh, và micrô trong môi trường âm thanh (ví dụ, nơi ở, ô tô hoặc nơi làm việc của người dùng) trong đó âm thanh có thể lan truyền từ người dùng đến micrô, và môi trường âm thanh có thể bao gồm các vùng xác định trước. Theo

một số ví dụ, môi trường có thể bao gồm ít nhất các vùng sau: khu vực chuẩn bị thực phẩm; khu vực ăn uống; khu vực mở cửa không gian phòng khách; khu vực TV (bao gồm ghế dài TV) của không gian phòng khách; và v.v. Trong quá trình vận hành hệ thống, giả sử rằng người dùng đang ở thực tế trong một trong các vùng (“vùng của người dùng”) bất kỳ lúc nào, và vùng của người dùng có thể thỉnh thoảng thay đổi.

Theo một số ví dụ, micrô có thể không đồng bộ (ví dụ, được lấy mẫu số bằng cách sử dụng đồng hồ mẫu riêng) và được đặt ngẫu nhiên (hoặc ít nhất là không nằm ở các vị trí định trước, sắp xếp đối xứng, trên lưới, v.v.). Trong một số trường hợp, vùng của người dùng có thể được ước lượng qua phương pháp điều khiển dữ liệu được điều khiển bằng nhiều đặc tính mức cao có được, ít nhất một phần, từ ít nhất một trong số các bộ phát hiện từ đánh thức. Theo một số ví dụ, các đặc tính này (ví dụ, mức tin cậy và nhận được từ đánh thức) có thể mất rất ít băng thông và có thể được truyền (ví dụ, không đồng bộ) cho thiết bị đang áp dụng bộ phân loại với rất ít phụ tải mạng.

Các khía cạnh của một số phương án bao gồm bước áp dụng các thiết bị âm thanh thông minh, và/hoặc phối hợp các thiết bị âm thanh thông minh.

Các khía cạnh của một số phương án thực hiện bao gồm hệ thống được tạo cấu hình (ví dụ, được lập trình) để thực hiện một hoặc nhiều phương pháp hoặc các bước được mô tả của chúng, và phương tiện bất biến hữu hình đọc được bằng máy tính thực hiện lưu trữ bất biến dữ liệu (ví dụ, đĩa hoặc phương tiện lưu trữ hữu hình khác) lưu trữ mã để thực hiện (ví dụ, mã có thể thực thi được để thực hiện) một hoặc nhiều phương pháp hoặc các bước của chúng. Ví dụ, một số phương án được mô tả có thể là hoặc bao gồm bộ xử lý đa năng có thể lập trình được, bộ xử lý tín hiệu số, hoặc bộ vi xử lý, được lập trình bằng phần mềm hoặc firmware và/hoặc được tạo cấu hình theo cách khác để thực hiện bất kỳ trong số nhiều thao tác trên dữ liệu, bao gồm một hoặc nhiều phương pháp hoặc các bước được mô tả của chúng. Bộ xử lý đa năng như vậy có thể là hoặc bao gồm hệ thống máy tính bao gồm thiết bị đầu vào, bộ nhớ, và hệ thống phụ xử lý được lập trình (và/hoặc được tạo cấu hình theo cách khác) để thực hiện một hoặc nhiều phương pháp được mô tả (hoặc các bước của chúng) để đáp lại dữ liệu đã được xác nhận như vậy.

Theo một số phương án thực hiện, hệ thống điều khiển có thể được tạo cấu hình để áp dụng một hoặc nhiều phương pháp được mô tả ở đây, như một hoặc nhiều phương

pháp quản lý phiên âm thanh. Một số phương pháp như vậy bao gồm bước nhận tín hiệu đầu ra (ví dụ, bằng hệ thống điều khiển) từ mỗi micrô trong số nhiều micrô trong môi trường âm thanh. Theo một số ví dụ, mỗi micrô trong số nhiều micrô nằm ở một vị trí micrô của môi trường âm thanh. Trong một số trường hợp, các tín hiệu đầu ra bao gồm các tín hiệu tương ứng với tiếng nói hiện tại của người. Theo một số ví dụ, các tín hiệu đầu ra bao gồm các tín hiệu tương ứng với dữ liệu âm thanh không phải lời nói, như tiếng ồn và/hoặc tiếng vọng.

Một số phương pháp như vậy bao gồm bước xác định (ví dụ, bằng hệ thống điều khiển), dựa trên các tín hiệu đầu ra, một hoặc nhiều khía cạnh của thông tin ngữ cảnh liên quan đến con người. Theo một số ví dụ, thông tin ngữ cảnh bao gồm vị trí hiện tại được ước lượng của người đó và/hoặc độ gần hiện tại được ước lượng của người đó tới một hoặc nhiều vị trí micrô. Một số phương pháp như vậy bao gồm bước lựa chọn hai hoặc nhiều thiết bị âm thanh của môi trường âm thanh dựa, ít nhất một phần, vào một hoặc nhiều khía cạnh của thông tin ngữ cảnh. Theo một số phương án thực hiện, mỗi trong số hai hoặc nhiều thiết bị âm thanh bao gồm ít nhất một loa.

Một số phương pháp như vậy bao gồm bước xác định (ví dụ, bằng hệ thống điều khiển) một hoặc nhiều kiểu thay đổi xử lý âm thanh để áp dụng cho dữ liệu âm thanh được kết xuất thành tín hiệu cấp cho loa cho hai hoặc nhiều thiết bị âm thanh. Theo một số ví dụ, các thay đổi xử lý âm thanh có tác dụng tăng tỷ lệ lời nói trên tiếng vọng tại một hoặc nhiều micrô. Một số phương pháp như vậy bao gồm bước khiến cho một hoặc nhiều kiểu thay đổi xử lý âm thanh sẽ được áp dụng.

Theo một số phương án thực hiện, một hoặc nhiều kiểu thay đổi xử lý âm thanh có thể khiến giảm mức tái tạo loa cho các loa của hai hoặc nhiều thiết bị âm thanh. Theo một số phương án thực hiện, ít nhất một trong số các thay đổi xử lý âm thanh đối với thiết bị âm thanh thứ nhất có thể khác với thay đổi xử lý âm thanh đối với thiết bị âm thanh thứ hai. Theo một số ví dụ, bước lựa chọn hai hoặc nhiều thiết bị âm thanh của môi trường âm thanh (ví dụ, bằng hệ thống điều khiển) có thể bao gồm bước lựa chọn N thiết bị âm thanh được trang bị loa của môi trường âm thanh, N là số nguyên lớn hơn 2.

Theo một số phương án thực hiện, bước lựa chọn hai hoặc nhiều thiết bị âm thanh của môi trường âm thanh có thể dựa, ít nhất một phần, vào vị trí hiện tại được ước lượng

của người so với ít nhất một trong số vị trí micrô hoặc vị trí thiết bị âm thanh được trang bị loa. Theo một số phương án thực hiện như vậy, phương pháp có thể bao gồm bước xác định thiết bị âm thanh được trang bị loa gần nhất tức là gần nhất với vị trí hiện tại được ước lượng của người đó hoặc với vị trí micrô gần nhất với vị trí hiện tại được ước lượng của người đó. Theo một số ví dụ như vậy, hai hoặc nhiều thiết bị âm thanh có thể bao gồm thiết bị âm thanh được trang bị loa gần nhất.

Theo một số ví dụ, một hoặc nhiều kiểu thay đổi xử lý âm thanh bao gồm bước thay đổi quy trình kết xuất để làm lệch việc kết xuất tín hiệu âm thanh khỏi vị trí hiện tại được ước lượng của người. Theo một số phương án thực hiện, một hoặc nhiều kiểu thay đổi xử lý âm thanh có thể bao gồm bước sửa đổi phổ. Theo một số phương án thực hiện như vậy, bước sửa đổi phổ có thể bao gồm bước giảm mức dữ liệu âm thanh trong băng tần số từ 500 Hz đến 3 KHz.

Theo một số phương án thực hiện, một hoặc nhiều kiểu thay đổi xử lý âm thanh có thể bao gồm bước chèn ít nhất một khe vào ít nhất một băng tần số đã chọn của tín hiệu phát lại âm thanh. Theo một số ví dụ, một hoặc nhiều kiểu thay đổi xử lý âm thanh có thể bao gồm nén dải động.

Theo một số phương án thực hiện, bước lựa chọn hai hoặc nhiều thiết bị âm thanh có thể dựa, ít nhất một phần, vào ước lượng tỷ lệ tín hiệu trên tiếng vọng cho một hoặc nhiều vị trí micrô. Ví dụ, bước lựa chọn hai hoặc nhiều thiết bị âm thanh có thể dựa, ít nhất một phần, vào bước xác định xem ước lượng tỷ lệ tín hiệu trên tiếng vọng có nhỏ hơn hoặc bằng ngưỡng tỷ lệ tín hiệu trên tiếng vọng không. Trong một số trường hợp, bước xác định một hoặc nhiều kiểu thay đổi xử lý âm thanh có thể dựa trên việc tối ưu hóa hàm chi phí dựa, ít nhất một phần, vào ước lượng tỷ lệ tín hiệu trên tiếng vọng. Ví dụ, hàm chi phí có thể dựa ít nhất một phần vào hiệu suất kết xuất. Theo một số phương án thực hiện, bước lựa chọn hai hoặc nhiều thiết bị âm thanh có thể dựa, ít nhất một phần, vào ước lượng độ gần.

Theo một số ví dụ, phương pháp này có thể bao gồm bước xác định (ví dụ, bằng hệ thống điều khiển) nhiều đặc tính âm thanh hiện tại từ tín hiệu đầu ra của mỗi micrô và áp dụng bộ phân loại cho nhiều đặc tính âm thanh hiện tại. Theo một số phương án thực hiện, bước áp dụng bộ phân loại có thể bao gồm bước áp dụng mô hình được huấn

luyện dựa trên các đặc tính âm thanh được xác định từ trước thu được từ nhiều tiếng nói trước đó của con người trong nhiều vùng người dùng trong môi trường.

Theo một số ví dụ như vậy, bước xác định một hoặc nhiều khía cạnh thông tin ngữ cảnh liên quan đến người có thể bao gồm bước xác định, dựa ít nhất một phần vào đầu ra từ bộ phân loại, ước lượng về vùng người dùng mà người đó hiện đang ở. Theo một số phương án thực hiện, việc ước lượng vùng người dùng có thể được xác định mà không cần tham chiếu đến các vị trí hình học của nhiều micrô. Trong một số trường hợp, tiếng nói hiện tại và các tiếng nói trước đó có thể là, hoặc có thể bao gồm, các tiếng nói từ đánh thức.

Theo một số phương án thực hiện, một hoặc nhiều micrô có thể nằm trong nhiều thiết bị âm thanh của môi trường âm thanh. Tuy nhiên, trong các trường hợp khác, một hoặc nhiều micrô có thể nằm trong một thiết bị âm thanh duy nhất của môi trường âm thanh. Theo một số ví dụ, ít nhất một trong số một hoặc nhiều vị trí micrô có thể tương ứng với nhiều micrô của một thiết bị âm thanh. Một số phương pháp được mô tả có thể bao gồm bước lựa chọn ít nhất một micrô theo một hoặc nhiều khía cạnh của thông tin ngữ cảnh.

Ít nhất một số khía cạnh của sáng chế có thể được thực hiện qua các phương pháp, như phương pháp quản lý phiên âm thanh. Như được lưu ý ở phần khác trong bản mô tả, trong một số trường hợp, các phương pháp có thể được thực hiện, ít nhất một phần, bằng hệ thống điều khiển như các phương pháp được mô tả ở đây. Một số phương pháp như vậy bao gồm bước nhận tín hiệu đầu ra từ mỗi micrô trong số nhiều micrô trong môi trường âm thanh. Theo một số ví dụ, mỗi micrô trong số nhiều micrô nằm ở vị trí micrô của môi trường âm thanh. Trong một số trường hợp, các tín hiệu đầu ra bao gồm các tín hiệu tương ứng với tiếng nói hiện tại của người. Theo một số ví dụ, các tín hiệu đầu ra bao gồm các tín hiệu tương ứng với dữ liệu âm thanh không phải lời nói, như tiếng ồn và/hoặc tiếng vọng.

Một số phương pháp như vậy bao gồm bước xác định, dựa trên các tín hiệu đầu ra, một hoặc nhiều khía cạnh của thông tin ngữ cảnh liên quan đến người. Theo một số ví dụ, thông tin ngữ cảnh bao gồm vị trí hiện tại được ước lượng của người đó và/hoặc độ gần hiện tại được ước lượng của người đó với một hoặc nhiều vị trí micrô. Một số phương pháp như vậy bao gồm bước lựa chọn hai hoặc nhiều thiết bị âm thanh của môi

trường âm thanh có thể dựa, ít nhất một phần, vào một hoặc nhiều khía cạnh của thông tin ngữ cảnh. Theo một số phương án thực hiện, mỗi trong số hai hoặc nhiều thiết bị âm thanh bao gồm ít nhất một loa.

Một số phương pháp như vậy bao gồm bước xác định một hoặc nhiều kiểu thay đổi xử lý âm thanh để áp dụng cho dữ liệu âm thanh được kết xuất thành tín hiệu cấp cho loa cho hai hoặc nhiều thiết bị âm thanh. Theo một số ví dụ, các thay đổi xử lý âm thanh có tác dụng tăng tỷ lệ lời nói trên tiếng vọng tại một hoặc nhiều micrô. Một số phương pháp như vậy bao gồm bước khiến cho một hoặc nhiều kiểu thay đổi xử lý âm thanh sẽ được áp dụng.

Theo một số phương án thực hiện, một hoặc nhiều kiểu thay đổi xử lý âm thanh có thể khiến giảm mức tái tạo loa cho các loa của hai hoặc nhiều thiết bị âm thanh. Theo một số phương án thực hiện, ít nhất một trong số các thay đổi xử lý âm thanh đối với thiết bị âm thanh thứ nhất có thể khác với thay đổi xử lý âm thanh đối với thiết bị âm thanh thứ hai. Theo một số ví dụ, bước lựa chọn hai hoặc nhiều thiết bị âm thanh của môi trường âm thanh có thể bao gồm bước lựa chọn N thiết bị âm thanh được trang bị loa của môi trường âm thanh, N là số nguyên lớn hơn 2.

Theo một số phương án thực hiện, bước lựa chọn hai hoặc nhiều thiết bị âm thanh của môi trường âm thanh có thể dựa, ít nhất một phần, vào vị trí hiện tại được ước lượng của người so với ít nhất một trong số vị trí micrô hoặc vị trí thiết bị âm thanh được trang bị loa. Theo một số phương án thực hiện như vậy, phương pháp có thể bao gồm bước xác định thiết bị âm thanh được trang bị loa gần nhất tức là gần nhất với vị trí hiện tại được ước lượng của người đó hoặc với vị trí micrô gần nhất với vị trí hiện tại được ước lượng của người đó. Theo một số ví dụ như vậy, hai hoặc nhiều thiết bị âm thanh có thể bao gồm thiết bị âm thanh được trang bị loa gần nhất.

Theo một số ví dụ, một hoặc nhiều kiểu thay đổi xử lý âm thanh bao gồm bước thay đổi quy trình kết xuất để làm lệch việc kết xuất tín hiệu âm thanh khỏi vị trí hiện tại được ước lượng của người. Theo một số phương án thực hiện, một hoặc nhiều kiểu thay đổi xử lý âm thanh có thể bao gồm bước sửa đổi phổ. Theo một số phương án thực hiện như vậy, bước sửa đổi phổ có thể bao gồm bước giảm mức dữ liệu âm thanh trong băng tần số từ 500 Hz đến 3 KHz.

Theo một số phương án thực hiện, một hoặc nhiều kiểu thay đổi xử lý âm thanh có thể bao gồm bước chèn ít nhất một khe vào ít nhất một băng tần số đã chọn của tín hiệu phát lại âm thanh. Theo một số ví dụ, một hoặc nhiều kiểu thay đổi xử lý âm thanh có thể bao gồm nén dải động.

Theo một số phương án thực hiện, bước lựa chọn hai hoặc nhiều thiết bị âm thanh có thể dựa, ít nhất một phần, vào ước lượng tỷ lệ tín hiệu trên tiếng vọng cho một hoặc nhiều vị trí micrô. Ví dụ, bước lựa chọn hai hoặc nhiều thiết bị âm thanh có thể dựa, ít nhất một phần, vào bước xác định xem ước lượng tỷ lệ tín hiệu trên tiếng vọng có nhỏ hơn hoặc bằng ngưỡng tỷ lệ tín hiệu trên tiếng vọng không. Trong một số trường hợp, bước xác định một hoặc nhiều kiểu thay đổi xử lý âm thanh có thể dựa trên việc tối ưu hóa hàm chi phí dựa, ít nhất một phần, vào ước lượng tỷ lệ tín hiệu trên tiếng vọng. Ví dụ, hàm chi phí có thể được dựa ít nhất một phần vào hiệu suất kết xuất. Theo một số phương án thực hiện, bước lựa chọn hai hoặc nhiều thiết bị âm thanh có thể được dựa, ít nhất một phần, vào ước lượng độ gần.

Theo một số ví dụ, phương pháp có thể bao gồm bước xác định nhiều đặc tính âm thanh hiện tại từ tín hiệu đầu ra của mỗi micrô và áp dụng bộ phân loại cho nhiều đặc tính âm thanh hiện tại. Theo một số phương án thực hiện, bước áp dụng bộ phân loại có thể bao gồm bước áp dụng mô hình được huấn luyện trên các đặc tính âm thanh được xác định từ trước thu được từ nhiều tiếng nói trước đó của người trong nhiều vùng người dùng trong môi trường.

Theo một số ví dụ như vậy, bước xác định một hoặc nhiều khía cạnh thông tin ngữ cảnh liên quan đến người có thể bao gồm bước xác định, dựa ít nhất một phần vào đầu ra từ bộ phân loại, ước lượng về vùng người dùng mà người đó hiện đang ở. Theo một số phương án thực hiện, ước lượng của vùng người dùng có thể được xác định mà không cần tham chiếu đến các vị trí hình học của nhiều micrô. Trong một số trường hợp, tiếng nói hiện tại và các tiếng nói trước đó có thể là, hoặc có thể bao gồm, các tiếng nói từ đánh thức.

Theo một số phương án thực hiện, một hoặc nhiều micrô có thể nằm trong nhiều thiết bị âm thanh của môi trường âm thanh. Tuy nhiên, trong các trường hợp khác, một hoặc nhiều micrô có thể nằm trong một thiết bị âm thanh duy nhất của môi trường âm thanh. Theo một số ví dụ, ít nhất một trong số một hoặc nhiều vị trí micrô có thể tương

ứng với nhiều micrô của một thiết bị âm thanh. Một số phương pháp được mô tả có thể bao gồm bước lựa chọn ít nhất một micrô theo một hoặc nhiều khía cạnh của thông tin ngữ cảnh.

Một số hoặc tất cả các hoạt động, chức năng và/hoặc phương pháp được mô tả ở đây có thể được thực hiện bằng một hoặc nhiều thiết bị theo các lệnh (ví dụ, phần mềm) được lưu trữ trên một hoặc nhiều phương tiện bất biến. Phương tiện bất biến như vậy có thể bao gồm các thiết bị nhớ như các thiết bị được mô tả ở đây, bao gồm nhưng không giới hạn ở thiết bị nhớ truy cập ngẫu nhiên (random access memory – RAM), thiết bị nhớ chỉ đọc (read-only memory – ROM), v.v. Do đó, một số khía cạnh sáng tạo của sáng chế được mô tả trong bản mô tả này có thể được thực hiện trong phương tiện bất biến có phần mềm được lưu trữ trên đó.

Ví dụ, phần mềm có thể bao gồm các lệnh để điều khiển một hoặc nhiều thiết bị thực hiện phương pháp bao gồm bước nhận tín hiệu đầu ra từ mỗi micrô trong số nhiều micrô trong môi trường âm thanh. Theo một số ví dụ, mỗi micrô trong số nhiều micrô nằm ở vị trí micrô của môi trường âm thanh. Trong một số trường hợp, các tín hiệu đầu ra bao gồm các tín hiệu tương ứng với tiếng nói hiện tại của người. Theo một số ví dụ, các tín hiệu đầu ra bao gồm các tín hiệu tương ứng với dữ liệu âm thanh không phải lời nói, như tiếng ồn và/hoặc tiếng vọng.

Một số phương pháp như vậy bao gồm bước xác định, dựa trên các tín hiệu đầu ra, một hoặc nhiều khía cạnh của thông tin ngữ cảnh liên quan đến người. Theo một số ví dụ, thông tin ngữ cảnh bao gồm vị trí hiện tại được ước lượng của người đó và/hoặc độ gần hiện tại được ước lượng của người đó với một hoặc nhiều vị trí micrô. Một số phương pháp như vậy bao gồm bước lựa chọn hai hoặc nhiều thiết bị âm thanh của môi trường âm thanh có thể dựa, ít nhất một phần, vào một hoặc nhiều khía cạnh của thông tin ngữ cảnh. Theo một số phương án thực hiện, mỗi trong số hai hoặc nhiều thiết bị âm thanh bao gồm ít nhất một loa.

Một số phương pháp như vậy bao gồm bước xác định một hoặc nhiều kiểu thay đổi xử lý âm thanh để áp dụng cho dữ liệu âm thanh được kết xuất thành tín hiệu cấp cho loa cho hai hoặc nhiều thiết bị âm thanh. Theo một số ví dụ, các thay đổi xử lý âm thanh có tác dụng tăng tỷ lệ lời nói trên tiếng vọng tại một hoặc nhiều micrô. Một số

phương pháp như vậy bao gồm bước khiến cho một hoặc nhiều kiểu thay đổi xử lý âm thanh sẽ được áp dụng.

Theo một số phương án thực hiện, một hoặc nhiều kiểu thay đổi xử lý âm thanh có thể khiến giảm mức tái tạo loa cho các loa của hai hoặc nhiều thiết bị âm thanh. Theo một số phương án thực hiện, ít nhất một trong số các thay đổi xử lý âm thanh đối với thiết bị âm thanh thứ nhất có thể khác với thay đổi xử lý âm thanh đối với thiết bị âm thanh thứ hai. Theo một số ví dụ, bước lựa chọn hai hoặc nhiều thiết bị âm thanh của môi trường âm thanh có thể bao gồm bước lựa chọn N thiết bị âm thanh được trang bị loa của môi trường âm thanh, N là số nguyên lớn hơn 2.

Theo một số phương án thực hiện, bước lựa chọn hai hoặc nhiều thiết bị âm thanh của môi trường âm thanh có thể dựa, ít nhất một phần, vào vị trí hiện tại được ước lượng của người so với ít nhất một trong số vị trí micrô hoặc vị trí thiết bị âm thanh được trang bị loa. Theo một số phương án thực hiện như vậy, phương pháp có thể bao gồm bước xác định thiết bị âm thanh được trang bị loa gần nhất tức là gần nhất với vị trí hiện tại được ước lượng của người đó hoặc với vị trí micrô gần nhất với vị trí hiện tại được ước lượng của người đó. Theo một số ví dụ như vậy, hai hoặc nhiều thiết bị âm thanh có thể bao gồm thiết bị âm thanh được trang bị loa gần nhất.

Theo một số ví dụ, một hoặc nhiều kiểu thay đổi xử lý âm thanh bao gồm bước thay đổi quy trình kết xuất để làm lệch việc kết xuất tín hiệu âm thanh khỏi vị trí hiện tại được ước lượng của người. Theo một số phương án thực hiện, một hoặc nhiều kiểu thay đổi xử lý âm thanh có thể bao gồm bước sửa đổi phổ. Theo một số phương án thực hiện như vậy, bước sửa đổi phổ có thể bao gồm bước giảm mức dữ liệu âm thanh trong băng tần số từ 500 Hz đến 3 KHz.

Theo một số phương án thực hiện, một hoặc nhiều kiểu thay đổi xử lý âm thanh có thể bao gồm bước chèn ít nhất một khe vào ít nhất một băng tần số đã chọn của tín hiệu phát lại âm thanh. Theo một số ví dụ, một hoặc nhiều kiểu thay đổi xử lý âm thanh có thể bao gồm nén dải động.

Theo một số phương án thực hiện, bước lựa chọn hai hoặc nhiều thiết bị âm thanh có thể dựa, ít nhất một phần, vào ước lượng tỷ lệ tín hiệu trên tiếng vọng cho một hoặc nhiều vị trí micrô. Ví dụ, bước lựa chọn hai hoặc nhiều thiết bị âm thanh có thể dựa, ít nhất một phần, vào bước xác định xem ước lượng tỷ lệ tín hiệu trên tiếng vọng có nhỏ

hơn hoặc bằng ngưỡng tỷ lệ tín hiệu trên tiếng vọng không. Trong một số trường hợp, bước xác định một hoặc nhiều kiểu thay đổi xử lý âm thanh có thể dựa trên việc tối ưu hóa hàm chi phí dựa, ít nhất một phần, vào ước lượng tỷ lệ tín hiệu trên tiếng vọng. Ví dụ, hàm chi phí có thể dựa ít nhất một phần vào hiệu suất kết xuất. Theo một số phương án thực hiện, bước lựa chọn hai hoặc nhiều thiết bị âm thanh có thể dựa, ít nhất một phần, vào ước lượng độ gần.

Theo một số ví dụ, phương pháp có thể bao gồm bước xác định nhiều đặc tính âm thanh hiện tại từ tín hiệu đầu ra của mỗi micrô và áp dụng bộ phân loại cho nhiều đặc tính âm thanh hiện tại. Theo một số phương án thực hiện, bước áp dụng bộ phân loại có thể bao gồm bước áp dụng mô hình được huấn luyện trên các đặc tính âm thanh được xác định từ trước thu được từ nhiều tiếng nói trước đó của con người trong nhiều vùng người dùng trong môi trường.

Theo một số ví dụ như vậy, bước xác định một hoặc nhiều khía cạnh thông tin ngữ cảnh liên quan đến người có thể bao gồm bước xác định, dựa ít nhất một phần vào đầu ra từ bộ phân loại, ước lượng về vùng người dùng mà người đó hiện đang ở. Theo một số phương án thực hiện, ước lượng của vùng người dùng có thể được xác định mà không cần tham chiếu đến các vị trí hình học của nhiều micrô. Trong một số trường hợp, tiếng nói hiện tại và các tiếng nói trước đó có thể là, hoặc có thể bao gồm, các tiếng nói từ đánh thức.

Theo một số phương án thực hiện, một hoặc nhiều micrô có thể nằm trong nhiều thiết bị âm thanh của môi trường âm thanh. Tuy nhiên, trong các trường hợp khác, một hoặc nhiều micrô có thể nằm trong một thiết bị âm thanh duy nhất của môi trường âm thanh. Theo một số ví dụ, ít nhất một trong số một hoặc nhiều vị trí micrô có thể tương ứng với nhiều micrô của một thiết bị âm thanh. Một số phương pháp được mô tả có thể bao gồm bước lựa chọn ít nhất một micrô theo một hoặc nhiều khía cạnh của thông tin ngữ cảnh.

Chi tiết của một hoặc nhiều phương án thực hiện của sáng chế được mô tả trong bản mô tả này được trình bày trong các hình vẽ kèm theo và phần mô tả dưới đây. Các đặc điểm, khía cạnh và ưu điểm khác sẽ trở nên rõ ràng nhờ phần mô tả, các hình vẽ, và phân yêu cầu bảo hộ. Lưu ý rằng kích thước tương đối của các hình vẽ sau đây không được vẽ theo tỷ lệ.

Mô tả vắn tắt các hình vẽ

Fig.1A thể hiện môi trường âm thanh theo một ví dụ.

Fig.1B thể hiện một ví dụ khác về môi trường âm thanh.

Fig.2A là sơ đồ khối thể hiện các ví dụ về các thành phần của thiết bị có khả năng thực hiện các khía cạnh khác nhau của sáng chế.

Fig.2B là lưu đồ bao gồm các khối của phương pháp quản lý phiên âm thanh theo một số phương án thực hiện.

Fig.3A là sơ đồ khối của hệ thống được tạo cấu hình để thực hiện điều khiển kết xuất và logic nghe hoặc thu nhận riêng trên nhiều thiết bị.

Fig.3B là sơ đồ khối của hệ thống theo phương án thực hiện khác được mô tả.

Fig.3C là sơ đồ khối của phương án được tạo cấu hình để áp dụng mạng cân bằng năng lượng theo một ví dụ.

Fig.4 là biểu đồ minh họa các ví dụ về xử lý âm thanh mà có thể làm tăng tỷ lệ lời nói trên tiếng vọng tại một hoặc nhiều micrô của môi trường âm thanh.

Fig.5 là biểu đồ minh họa kiểu xử lý âm thanh khác mà có thể làm tăng tỷ lệ lời nói trên tiếng vọng tại một hoặc nhiều micrô của môi trường âm thanh.

Fig.6 minh họa kiểu xử lý âm thanh khác mà có thể làm tăng tỷ lệ lời nói trên tiếng vọng tại một hoặc nhiều micrô của môi trường âm thanh.

Fig.7 là biểu đồ minh họa kiểu xử lý âm thanh khác mà có thể làm tăng tỷ lệ lời nói trên tiếng vọng tại một hoặc nhiều micrô của môi trường âm thanh.

Fig.8 là sơ đồ của một ví dụ trong đó thiết bị âm thanh được giảm âm lượng có thể không phải là thiết bị âm thanh gần nhất với người đang nói.

Fig.9 minh họa một tình huống trong đó thiết bị có SER rất cao ở rất gần người dùng.

Fig.10 là lưu đồ mô tả ví dụ về phương pháp có thể được thực hiện bằng thiết bị như thiết bị được thể hiện trên Fig.2A.

Fig.11 là sơ đồ khối của các phần tử theo một phương án làm ví dụ được tạo cấu hình để triển khai bộ phân loại vùng.

Fig.12 là lưu đồ mô tả phương pháp làm ví dụ có thể được thực hiện bằng thiết bị như thiết bị 200 trên Fig.2A.

Fig.13 là lưu đồ mô tả một phương pháp làm ví dụ khác có thể được thực hiện bằng một thiết bị như thiết bị 200 trên Fig.2A.

Fig.14 là lưu đồ mô tả một phương pháp làm ví dụ khác có thể được thực hiện bằng thiết bị như thiết bị 200 trên Fig.2A.

Fig.15 và Fig.16 là sơ đồ minh họa ví dụ về tập hợp kích hoạt loa và các vị trí kết xuất đối tượng.

Fig.17 là lưu đồ mô tả phương pháp làm ví dụ có thể được thực hiện bằng thiết bị hoặc hệ thống như thể hiện trên Fig.2A.

Fig.18 là biểu đồ kích hoạt loa theo một phương án làm ví dụ.

Fig.19 là biểu đồ các vị trí kết xuất đối tượng theo một phương án làm ví dụ.

Fig.20 là biểu đồ kích hoạt loa theo một phương án làm ví dụ.

Fig.21 là biểu đồ các vị trí kết xuất đối tượng theo một phương án làm ví dụ.

Fig.22 là biểu đồ kích hoạt loa theo một phương án làm ví dụ.

Fig.23 là biểu đồ các vị trí kết xuất đối tượng theo một phương án làm ví dụ.

Mô tả chi tiết sáng chế

Hiện nay, các nhà thiết kế thường coi các thiết bị âm thanh như một điểm giao diện của âm thanh mà có thể là sự kết hợp của các dịch vụ giải trí, truyền thông và thông tin. Việc sử dụng âm thanh để thông báo và điều khiển bằng giọng nói có ưu điểm là tránh được việc xâm nhập trực quan hoặc vật lý. Bối cảnh thiết bị âm thanh mở rộng đang ngày càng phân hóa đa dạng, với nhiều hệ thống hơn đang cạnh tranh để phục vụ đôi tai con người.

Trong tất cả các dạng âm thanh tương tác, vấn đề cải thiện khả năng âm thanh song công toàn phần vẫn là một thách thức. Khi có đầu ra âm thanh trong phòng không thích hợp để truyền hoặc thu dựa trên thông tin trong phòng, mong muốn là loại bỏ âm thanh này khỏi tín hiệu đã thu (ví dụ, bằng cách loại bỏ tiếng vọng và/hoặc khử tiếng vọng). Một số phương án được mô tả đề xuất phương pháp tiếp cận và quản lý trải nghiệm người dùng để cải thiện tỷ lệ Tín hiệu trên tiếng vọng (Signal to Echo ratio – SER), đây là tiêu chí chính để song công toàn phần thành công ở một hoặc nhiều thiết bị.

Các phương án như vậy được kỳ vọng là có ích trong các tình huống có nhiều hơn một thiết bị âm thanh nằm trong phạm vi âm thanh của người dùng, sao cho mỗi

thiết bị âm thanh có khả năng biểu diễn nội dung chương trình âm thanh có âm lượng lớn thích hợp cho người dùng đối với dịch vụ giải trí, truyền thông hoặc thông tin mong muốn. Giá trị của các phương án như vậy được dự kiến đặc biệt cao khi có ba hoặc nhiều thiết bị âm thanh ở gần người dùng tương đương nhau.

Ứng dụng kết xuất đôi khi là chức năng chính của thiết bị âm thanh, và vì vậy, đôi khi có mong muốn sử dụng càng nhiều thiết bị đầu ra âm thanh càng tốt. Nếu thiết bị âm thanh gần người dùng hơn, thì thiết bị âm thanh có thể có lợi thế hơn về khả năng định vị âm thanh chính xác hoặc cung cấp tín hiệu âm thanh và hình ảnh cụ thể cho người dùng. Tuy nhiên, nếu các thiết bị âm thanh bao gồm một hoặc nhiều micrô, thì chúng cũng có thể thích hợp hơn để thu giọng nói của người dùng. Khi xem xét thách thức về tỷ lệ tín hiệu trên tiếng vọng, chúng tôi thấy rằng tỷ lệ tín hiệu trên tiếng vọng được cải thiện đáng kể nếu chúng tôi áp dụng việc sử dụng thiết bị gần với người dùng hơn ở chế độ đơn công (chỉ nhập) hoặc chuyển sang chế độ này.

Theo một số phương án được mô tả khác nhau, thiết bị âm thanh có thể có cả loa và micrô tích hợp trong khi phục vụ các chức năng khác (ví dụ, các chức năng được thể hiện trên Fig.1A). Một số phương án được mô tả áp dụng khái niệm chủ động không sử dụng (các) loa gần nhất với người dùng trong một số trường hợp.

Dự kiến rằng trong hệ điều hành được kết nối hoặc phi trung gian giữa các ứng dụng (ví dụ, các ứng dụng dựa trên đám mây), các thiết bị thuộc nhiều loại khác nhau (cho phép đầu vào, đầu ra âm thanh và/hoặc tương tác thời gian thực) có thể được bao gồm. Ví dụ về các thiết bị như vậy bao gồm thiết bị đeo được, thiết bị âm thanh gia đình, thiết bị di động, thiết bị tính toán di động và trên ô tô, và loa thông minh. Loa thông minh có thể bao gồm loa và micrô được kết nối mạng cho các dịch vụ dựa trên đám mây. Các ví dụ khác về các thiết bị như vậy có thể tích hợp loa và/hoặc micrô, bao gồm đèn, đồng hồ, TV, thiết bị hỗ trợ cá nhân, tủ lạnh và thùng rác. Một số phương án có liên quan đặc biệt đến các tình huống trong đó có nền tảng chung để phối nhạc cho nhiều thiết bị âm thanh của môi trường âm thanh qua thiết bị hòa âm như trung tâm nhà thông minh hoặc thiết bị khác được tạo cấu hình để quản lý phiên âm thanh, ở đây có thể được gọi là bộ quản lý phiên âm thanh. Một số phương án thực hiện như vậy có thể liên quan đến các lệnh giữa bộ quản lý phiên âm thanh và các ứng dụng phần mềm được thực hiện cục bộ theo ngôn ngữ không dành riêng cho thiết bị, nhưng thay vào đó bao gồm thiết

bị hòa âm định tuyến nội dung âm thanh đến và đi từ người và địa điểm được xác định bằng ứng dụng phần mềm. Một số phương án áp dụng các phương pháp để quản lý kết xuất động, ví dụ, bao gồm ràng buộc để đẩy âm thanh ra khỏi thiết bị gần nhất và duy trì hình ảnh không gian, và/hoặc để định vị người dùng trong một vùng, và/hoặc lập bản đồ và vị trí của các thiết bị liên quan với nhau và người dùng.

Thông thường, hệ thống bao gồm nhiều thiết bị âm thanh thông minh cần thiết để biểu thị khi nào nó đã nghe thấy “tức đánh thức” (được định nghĩa ở trên) từ người dùng và chú ý đến (nói cách khác, nghe) lệnh từ người dùng.

Fig.1A thể hiện môi trường âm thanh theo một ví dụ. Một số phương án được mô tả có thể đặc biệt hữu ích trong trường hợp có nhiều thiết bị âm thanh trong môi trường (ví dụ, không gian phòng khách hoặc phòng làm việc) mà có thể truyền âm thanh và ghi lại âm thanh, như được mô tả ở đây. Hệ thống trên Fig.1A có thể được tạo cấu hình theo một số phương án được mô tả.

Fig.1A là sơ đồ của môi trường âm thanh (không gian phòng khách) bao gồm hệ thống gồm tập hợp thiết bị âm thanh thông minh (thiết bị 1.1) để tương tác âm thanh, loa (1.3) để xuất ra âm thanh, và đèn điều khiển được (1.2). Cũng như các phương án thực hiện được mô tả khác, kiểu, số lượng và bố trí các phần tử trên Fig.1A chỉ là ví dụ. Các phương án thực hiện khác có thể cung cấp nhiều phần tử hơn, ít phần tử hơn và/hoặc các phần tử khác nhau. Trong một số trường hợp, một hoặc nhiều micro 1.5 có thể là một phần của, hoặc kết hợp với một trong các thiết bị 1.1, đèn 1.2 hoặc loa 1.3. Cách khác, hoặc ngoài ra, một hoặc nhiều micro 1.5 có thể được gắn vào một phần khác của môi trường, ví dụ, vào tường, lên trần nhà, vào đồ nội thất, thiết bị hoặc vào thiết bị khác của môi trường. Theo một ví dụ, mỗi thiết bị 1.1 bao gồm (và/hoặc được kết hợp với) ít nhất một micro 1.5. Mặc dù không được thể hiện trên Fig.1A, một số môi trường âm thanh có thể bao gồm một hoặc nhiều camera. Theo một số phương án thực hiện được mô tả, một hoặc nhiều thiết bị của môi trường âm thanh (ví dụ, thiết bị được tạo cấu hình để quản lý phiên âm thanh, như một hoặc nhiều thiết bị 1.1, thiết bị mà đang áp dụng bộ quản lý phiên âm thanh, trung tâm (hub) nhà thông minh, v.v.) có thể ước lượng xem người dùng (1.4) đưa ra từ đánh thức, mệnh lệnh, v.v. ở đâu (ví dụ, ở vùng nào của không gian phòng khách). Một hoặc nhiều thiết bị của hệ thống thể hiện trên Fig.1A (ví dụ, các thiết bị 1.1 của chúng) có thể được tạo cấu hình để áp dụng các phương án được

mô tả khác nhau. Bằng cách sử dụng các phương pháp khác nhau, thông tin có thể được thu nhận chung từ các thiết bị trên Fig.3 để cung cấp ước lượng vị trí của người dùng đã nói từ đánh thức. Theo một số phương pháp được mô tả, thông tin có thể được thu thập chung từ micro 1.5 trên Fig.1A và được cung cấp cho thiết bị (ví dụ, thiết bị được tạo cấu hình để quản lý phiên âm thanh), áp dụng bộ phân loại được tạo cấu hình để cung cấp ước lượng vị trí của người dùng nói từ đánh thức.

Trong không gian phòng khách (ví dụ, trên Fig.1A), có một tập hợp các vùng hoạt động tự nhiên, nơi con người sẽ thực hiện một nhiệm vụ hoặc hoạt động, hoặc vượt qua một ngưỡng. Theo một số ví dụ, những khu vực này, có thể được gọi ở đây là vùng người dùng, có thể được người dùng xác định, mà không xác định tọa độ hoặc chỉ dẫn khác về vị trí hình học. Theo một số ví dụ, “ngữ cảnh” của người có thể bao gồm, hoặc có thể tương ứng với, vùng người dùng mà người đó hiện đang ở hoặc ước lượng về vùng đó. Trong ví dụ trên Fig.1A, vùng người dùng bao gồm:

1. Khu vực bồn rửa bếp và chuẩn bị thức ăn (ở vùng phía trên bên trái của không gian phòng khách);
2. Cửa tủ lạnh (bên phải bồn rửa và khu vực chuẩn bị thức ăn);
3. Khu ăn uống (ở vùng dưới bên trái của không gian phòng khách);
4. Khu vực mở cửa không gian phòng khách (bên phải khu vực bồn rửa và chuẩn bị thức ăn và khu ăn uống);
5. Ghế dài ti vi (TV) (ở bên phải của khu vực mở);
6. TV;
7. Bàn; và
8. Khu vực cửa ra vào hoặc lối vào (ở vùng trên bên phải của không gian phòng khách). Các môi trường âm thanh khác có thể bao gồm nhiều hơn, ít hơn và/hoặc các loại vùng người dùng khác, như một hoặc nhiều khu phòng ngủ, khu ga ra, khu sân trong hoặc sân ngoài, v.v.

Theo một số phương án, hệ thống ước lượng (ví dụ, xác định ước lượng không chắc chắn về) nơi phát ra hoặc bắt nguồn âm thanh (ví dụ, từ đánh thức hoặc tín hiệu gây chú ý khác) có thể có độ tin cậy xác định nào đó về (hoặc nhiều giả thuyết cho) ước lượng. Ví dụ, nếu người tình cờ ở gần biên giữa các vùng người dùng của môi trường âm thanh, thì việc ước lượng không chắc chắn về vị trí của người đó có thể bao gồm độ

tin cậy được xác định mà người đó đang ở trong mỗi vùng. Theo một số phương án thực hiện thông thường của giao diện giọng nói, giọng nói của trợ lý giọng nói chỉ được phát ra từ một vị trí tại một thời điểm, do đó buộc phải có một lựa chọn duy nhất cho một vị trí duy nhất (ví dụ, một trong tám vị trí loa, 1.1 và 1.3, trên Fig.1A). Tuy nhiên, dựa trên việc sắm vai tưởng tượng đơn giản, rõ ràng là (trong các phương án thực hiện thông thường như vậy) khả năng của vị trí được lựa chọn của nguồn của trợ lý giọng nói (nghĩa là, vị trí của loa được bao gồm trong hoặc kết hợp với trợ lý) là điểm tập trung hoặc phản hồi trả về tự nhiên để thể hiện sự chú ý có thể thấp.

Fig.1B thể hiện một ví dụ khác về môi trường âm thanh. Fig.1B mô tả môi trường âm thanh khác bao gồm người dùng 101 phát ra lời nói trực tiếp 102, và hệ thống bao gồm tập hợp các thiết bị âm thanh thông minh 103 và 105, loa cho đầu ra âm thanh, và các micro. Hệ thống có thể được tạo cấu hình theo một số phương án thực hiện được mô tả. Lời nói được phát ra bởi người dùng 101 (đôi khi được gọi ở đây là người nói) có thể được một hoặc nhiều phần tử của hệ thống nhận dạng là từ đánh thức.

Cụ thể hơn, các phần tử của hệ thống trên Fig.1B bao gồm:

102: giọng nói cục bộ trực tiếp (do người dùng 101 tạo ra);

103: thiết bị trợ lý giọng nói (gắn với một hoặc nhiều loa). Thiết bị 103 được đặt gần người dùng 101 hơn là thiết bị 105, và do đó thiết bị 103 đôi khi được gọi là thiết bị “gần”, và thiết bị 105 được gọi là thiết bị “ở xa”;

104: nhiều micro trong (hoặc được gắn với) thiết bị gần 103;

105: thiết bị trợ lý giọng nói (được gắn với một hoặc nhiều loa);

106: nhiều micro trong (hoặc được gắn với) thiết bị ở xa 105;

107: Thiết bị gia dụng (ví dụ, đèn); và

108: Nhiều micro trong (hoặc được gắn với) thiết bị gia dụng 107. Theo một số ví dụ, mỗi trong số các micro 108 có thể được tạo cấu hình để truyền thông với thiết bị được tạo cấu hình để áp dụng bộ phân loại, trong một số trường hợp, có thể có ít nhất một trong các thiết bị 103 hoặc 105. Theo một số phương án thực hiện, thiết bị được tạo cấu hình để áp dụng bộ phân loại cũng có thể là thiết bị được tạo cấu hình để quản lý phiên âm thanh, như thiết bị được tạo cấu hình để áp dụng CHASM hoặc trung tâm nhà thông minh.

Hệ thống trên Fig.1B cũng có thể bao gồm ít nhất một bộ phân loại (ví dụ, bộ phân loại 1107 trên Fig.11 được mô tả dưới đây). Ví dụ, thiết bị 103 (hoặc thiết bị 105) có thể bao gồm bộ phân loại. Cách khác, hoặc ngoài ra, bộ phân loại có thể được thực hiện bằng thiết bị khác có thể được tạo cấu hình để truyền thông với các thiết bị 103 và/hoặc 105. Theo một số ví dụ, bộ phân loại có thể được thực hiện bằng thiết bị cục bộ khác (ví dụ, thiết bị trong môi trường 109), trong khi trong các ví dụ khác, bộ phân loại có thể được thực hiện bằng thiết bị từ xa đặt bên ngoài môi trường 109 (ví dụ, máy chủ).

Theo một số phương án thực hiện, ít nhất hai thiết bị (ví dụ, thiết bị 1.1 trên Fig.1A, thiết bị 103 và 105 trên Fig.1B, v.v.) theo cách nào đó làm việc cùng nhau (ví dụ, dưới sự điều khiển của thiết bị hòa âm, như thiết bị được tạo cấu hình để quản lý phiên âm thanh) để cung cấp âm thanh trong đó âm thanh có thể được kiểm soát chung giữa chúng. Ví dụ, hai thiết bị 103 và 105 có thể phát âm thanh riêng hoặc chung. Trong trường hợp đơn giản, thiết bị 103 và 105 đóng vai trò như một cặp khớp nối đối với mỗi phần kết xuất của âm thanh (ví dụ, không làm mất tính tổng quát, tín hiệu âm thanh nổi có một phần kết xuất về cơ bản là L và phần còn lại về cơ bản là R).

Thiết bị 107 (hoặc thiết bị khác) có thể bao gồm một micrô 108 gần với người dùng 101 nhất và không bao gồm bất kỳ loa nào, trong trường hợp này, chúng ta có tình huống trong đó có thể đã có Tỷ lệ tín hiệu trên tiếng vọng hoặc Tỷ lệ giọng trên tiếng vọng (Speech to Echo Ratio – SER) thích hợp cho môi trường âm thanh cụ thể này và vị trí cụ thể này của người dùng 101 không thể được cải thiện bằng cách thay đổi quy trình xử lý âm thanh đối với âm thanh được tái tạo bằng (các) loa của thiết bị 105 và/hoặc 107. Theo một số phương án, không có micrô như vậy.

Một số phương án được mô tả đề xuất ảnh hưởng hiệu suất SER phát hiện được và quan trọng. Một số phương án thực hiện cung cấp những ưu điểm như vậy mà không cần áp dụng các khía cạnh của vị trí vùng và/hoặc kết xuất biến động. Tuy nhiên, một số phương án áp dụng các thay đổi xử lý âm thanh bao gồm bước kết xuất bằng cách đẩy lùi hoặc “làm lệch” các đối tượng âm thanh (hoặc vật thể âm thanh) ra khỏi thiết bị. Lý do làm lệch các đối tượng âm thanh từ các thiết bị âm thanh, vị trí cụ thể, v.v., trong một số trường hợp, có thể là để cải thiện Tỷ lệ tín hiệu trên tiếng vọng tại một micrô cụ thể được dùng để thu lời nói của con người. Việc làm lệch như vậy có thể liên quan đến, nhưng có thể không giới hạn ở, việc giảm mức âm lượng phát lại của một, hai, ba hoặc

nhiều thiết bị âm thanh gần đó. Trong một số trường hợp, các thay đổi trong xử lý âm thanh để cải thiện SER có thể được thông báo bằng kỹ thuật phát hiện vùng sao cho một, hai hoặc nhiều thiết bị âm thanh ở gần thực hiện các thay đổi trong xử lý âm thanh (ví dụ, được giảm âm lượng) là các thiết bị bị gần người dùng nhất, gần micro cụ thể nhất sẽ được dùng để thu lời nói của người dùng và/hoặc gần âm thanh quan tâm nhất.

Các khía cạnh của một số phương án liên quan đến ngữ cảnh, quyết định và thay đổi xử lý âm thanh, có thể được gọi ở đây là “thay đổi kết xuất”. Theo một số ví dụ, các khía cạnh này là:

NGỮ CẢNH (như vị trí và/hoặc thời gian). Theo một số ví dụ, cả vị trí và thời gian đều là một phần của ngữ cảnh, và mỗi vị trí và thời gian có thể được suy ra hoặc xác định theo một số cách khác nhau;

QUYẾT ĐỊNH (có thể bao gồm ngưỡng hoặc điều chế liên tục (các) thay đổi). Thành phần này có thể đơn giản hoặc phức tạp, tùy thuộc vào phương án thực hiện cụ thể. Theo một số phương án, quyết định có thể được đưa ra trên cơ sở liên tục, ví dụ, theo phản hồi. Trong một số trường hợp, quyết định có thể tạo ra sự ổn định của hệ thống, ví dụ, sự ổn định phản hồi tốt như được mô tả dưới đây; và

KẾT XUẤT (bản chất của (các) thay đổi xử lý âm thanh). Mặc dù được gọi ở đây là “kết xuất”, nhưng (các) thay đổi xử lý âm thanh có thể có hoặc không bao gồm (các) thay đổi kết xuất, tùy thuộc vào phương án thực hiện cụ thể. Theo một số phương án thực hiện, có một số tùy chọn để thay đổi xử lý âm thanh, bao gồm việc áp dụng các thay đổi xử lý âm thanh khó cảm nhận thông qua việc áp dụng kết xuất các thay đổi xử lý âm thanh đơn giản và hiển nhiên.

Theo một số ví dụ, “ngữ cảnh” có thể bao gồm thông tin về cả vị trí và mục đích. Ví dụ, thông tin ngữ cảnh có thể bao gồm ít nhất một ý tưởng sơ bộ về vị trí của người dùng, như ước lượng về vùng người dùng tương ứng với vị trí hiện tại của người dùng. Thông tin ngữ cảnh có thể tương ứng với vị trí đối tượng âm thanh, ví dụ, vị trí đối tượng âm thanh tương ứng với tiếng nói từ đánh thức của người dùng. Theo một số ví dụ, thông tin ngữ cảnh có thể bao gồm thông tin về thời gian và khả năng của đối tượng hoặc cá nhân đó phát ra âm thanh. Các ví dụ về ngữ cảnh bao gồm, nhưng không giới hạn, như sau:

A. Biết về vị trí có thể có. Việc này có thể dựa trên

i) phát hiện yếu hoặc xác suất thấp (ví dụ, phát hiện âm thanh mà có thể được quan tâm nhưng có thể đủ rõ ràng hoặc có thể không đủ rõ ràng để hành động);

ii) kích hoạt riêng (ví dụ, từ đánh thức được nói và được phát hiện rõ ràng);

iii) thói quen và các mẫu (ví dụ, dựa trên nhận dạng mẫu, ví dụ, một số vị trí nhất định như ghế dài gần TV có thể được kết hợp với việc một hoặc nhiều người đang xem nội dung video trên TV và nghe âm thanh liên quan khi ngồi trên ghế dài);

iv) và/hoặc tích hợp dạng cảm biến độ gần khác nào đó dựa trên phương thức khác (như một hoặc nhiều cảm biến hồng ngoại (infrared – IR), camera, cảm biến điện dung, cảm biến tần số vô tuyến (radio frequency – RF), cảm biến nhiệt, cảm biến áp suất) (ví dụ, trong hoặc trên đồ nội thất của môi trường âm thanh), đài radar đeo được, v.v.); và

B. Biết hoặc ước lượng khả năng để âm thanh mà một người có thể muốn nghe, ví dụ, có khả năng phát hiện được cải thiện. Việc này có thể bao gồm một số hoặc tất cả trong số:

i) sự kiện dựa trên việc phát hiện âm thanh nào đó như phát hiện từ đánh thức;

ii) sự kiện hoặc ngữ cảnh dựa trên hoạt động đã biết hoặc chuỗi sự kiện, ví dụ, tạm dừng hiển thị nội dung video, không gian tương tác trong nội dung tương tác kiểu nhận dạng lời nói tự động (automatic speech recognition – ASR) theo kịch bản, hoặc các thay đổi về hoạt động và/hoặc động lực và/hoặc hội thoại của hoạt động truyền thông song công toàn phần (như việc tạm dừng bởi một hoặc nhiều người tham gia hội nghị từ xa);

iii) đầu vào nhận biết bổ sung của các phương thức khác;

iv) lựa chọn để khả năng nghe liên tục được cải thiện theo cách nào đó – nâng cao tính sẵn sàng hoặc cải thiện khả năng nghe.

Điểm khác biệt cơ bản giữa A (biết vị trí có thể ở đâu) và B (biết hoặc ước lượng khả năng của âm thanh mà chúng ta muốn nghe, ví dụ, với khả năng phát hiện được cải thiện) là A có thể bao gồm thông tin hoặc kiến thức về vị trí cụ thể mà không nhất thiết phải biết xem có thông tin gì để nghe hay không, trong khi B có thể tập trung hơn vào

thông tin thời gian hoặc sự kiện cụ thể mà không nhất thiết phải biết chính xác nơi để nghe. Tất nhiên có thể có sự trùng hợp trong một số khía cạnh của A và B, ví dụ, việc phát hiện từ đánh thức yếu hoặc đầy đủ sẽ có thông tin về cả vị trí và thời gian.

Đối với việc sử dụng một số trường hợp, điều có thể quan trọng là “ngữ cảnh” bao gồm thông tin về cả vị trí (ví dụ, vị trí của người và/hoặc của micrô gần đó) và thời điểm muốn nghe. Thông tin theo ngữ cảnh này có thể thúc đẩy một hoặc nhiều quyết định liên quan và một hoặc nhiều thay đổi xử lý âm thanh có thể có (ví dụ, một hoặc nhiều thay đổi kết xuất có thể có). Do đó, một số phương án khác nhau cho phép sử dụng nhiều khả năng dựa trên các loại thông tin khác nhau mà có thể được dùng để tạo ra ngữ cảnh.

Trong phần tiếp theo, sáng chế mô tả khía cạnh “quyết định”. Ví dụ, khía cạnh này có thể bao gồm bước xác định một, hai, ba hoặc nhiều thiết bị đầu ra mà quy trình xử lý âm thanh liên quan sẽ được thay đổi. Một cách đơn giản để đưa ra quyết định như vậy là:

Cung cấp thông tin từ ngữ cảnh (ví dụ, vị trí và/hoặc sự kiện (hoặc theo một nghĩa nào đó là sự tin cậy về cái gì đó có ý nghĩa hoặc sự quan trọng về vị trí đó)), theo một số ví dụ, bộ quản lý phiên âm thanh có thể xác định hoặc ước lượng khoảng cách từ vị trí đó đến một số hoặc tất cả các thiết bị âm thanh trong môi trường âm thanh. Theo một số phương án thực hiện, bộ quản lý phiên âm thanh cũng có thể tạo ra tập hợp điện thế hoạt động cho mỗi loa (hoặc tập hợp loa) cho một số hoặc tất cả các thiết bị âm thanh của môi trường âm thanh. Theo một số ví dụ như vậy, tập hợp điện thế hoạt động có thể được xác định là $[f_1, f_2, \dots, f_n]$ và không làm mất tính tổng quát nằm trong khoảng $[0 \dots 1]$. Theo một ví dụ khác, kết quả của quyết định có thể mô tả việc cải thiện tỷ lệ lời nói trên tiếng vọng đích $[s_1, s_2, \dots, s_n]$ trên mỗi thiết bị theo khía cạnh “kết xuất”. Theo một ví dụ khác, cả điện thế hoạt động và việc cải thiện tỷ lệ giọng nói thành tiếng vọng đều có thể được tạo ra theo khía cạnh “quyết định”.

Theo một số phương án, điện thế hoạt động thông báo mức mà khía cạnh “kết xuất” phải đảm bảo rằng SER được cải thiện ở vị trí micrô mong muốn. Theo một số ví dụ như vậy, các giá trị lớn nhất f_n có thể biểu thị rằng âm thanh kết xuất được giảm (ducking) hoặc làm lệch đáng kể, hoặc trong trường hợp mà giá trị s_n được cung cấp, âm thanh bị hạn chế và giảm để đạt được tỷ lệ lời nói trên tiếng vọng s_n . Giá trị trung

biên f_n gần 0,5 hơn theo một số phương án có thể biểu thị rằng chỉ cần thay đổi kết xuất ở mức độ vừa phải, và có thể thích hợp để làm lệch các nguồn âm thanh cho các vị trí này. Hơn nữa, theo một số phương án thực hiện, các giá trị thấp f_n có thể được coi là không quan trọng để giảm bớt. Theo một số phương án thực hiện như vậy, các giá trị f_n bằng hoặc thấp hơn mức ngưỡng có thể không được xác nhận. Theo một số ví dụ, các giá trị f_n bằng hoặc thấp hơn mức ngưỡng có thể tương ứng với các vị trí làm lệch việc kết xuất nội dung âm thanh về một hướng. Trong một số trường hợp, các loa tương ứng với các giá trị f_n bằng hoặc thấp hơn mức ngưỡng thậm chí có thể được nâng lên ở mức phát lại theo một số quy trình được mô tả sau.

Theo một số phương án thực hiện, phương pháp nêu trên (hoặc một trong các phương pháp thay thế được mô tả dưới đây) có thể được dùng để tạo ra thông số điều khiển cho từng thay đổi xử lý âm thanh đã chọn cho tất cả các thiết bị âm thanh đã chọn, ví dụ, cho từng thiết bị của môi trường âm thanh, cho một hoặc nhiều thiết bị của môi trường âm thanh, cho hai hoặc nhiều thiết bị của môi trường âm thanh, cho ba hoặc nhiều thiết bị của môi trường âm thanh, v.v. Việc lựa chọn các thay đổi xử lý âm thanh có thể khác nhau theo phương án thực hiện cụ thể. Ví dụ, quyết định có thể bao gồm bước xác định:

- Tập hợp hai hoặc nhiều loa để thay đổi quy trình xử lý âm thanh; và
- Mức độ để thay đổi quy trình xử lý âm thanh cho tập hợp hai hoặc nhiều loa.

Theo một số ví dụ, mức độ thay đổi có thể được xác định trong ngữ cảnh của phạm vi được thiết kế hoặc được xác định, mà có thể dựa ít nhất một phần vào khả năng của một hoặc nhiều loa trong tập hợp loa. Trong một số trường hợp, khả năng của mỗi loa có thể bao gồm đáp ứng tần số, giới hạn mức phát lại và/hoặc các thông số của một hoặc nhiều thuật toán xử lý động của loa.

Ví dụ, trong một trường hợp cụ thể lựa chọn thiết kế có thể là lựa chọn tốt nhất là giảm âm lượng loa. Theo một số ví dụ như vậy, mức tối đa và/hoặc tối thiểu của thay đổi xử lý âm thanh có thể được xác định, ví dụ, mức độ mà mọi loa sẽ được giảm âm lượng được giới hạn ở một ngưỡng cụ thể, ví dụ, 15 dB, 20 dB, 25 dB, v.v. Theo một số phương án thực hiện như vậy, quyết định có thể dựa trên heuristic hoặc logic để lựa chọn một, hai, ba hoặc nhiều loa, và dựa trên mức tin cậy của hoạt động quan tâm, vị trí loa, v.v., quyết định có thể là giảm âm thanh được tái tạo bằng một, hai, ba hoặc nhiều

loa bằng một lượng trong phạm vi giá trị tối đa và tối thiểu, ví dụ, từ 0 đến 20 dB. Trong một số trường hợp, phương pháp (hoặc phần tử hệ thống) quyết định có thể tạo ra tập hợp điện thế hoạt động cho mỗi thiết bị âm thanh được trang bị loa.

Theo một ví dụ đơn giản, quy trình quyết định có thể đơn giản là xác định rằng tất cả các thiết bị âm thanh khác thiết bị âm thanh có thay đổi kích hoạt kết xuất bằng 0 và xác định rằng một thiết bị âm thanh có thay đổi kích hoạt bằng 1. Thiết kế của (các) thay đổi xử lý âm thanh (ví dụ, giảm) và mức độ của (các) thay đổi xử lý âm thanh (ví dụ, hằng số thời gian, v.v.), theo một số ví dụ, có thể độc lập với logic quyết định. Phương pháp này tạo ra một thiết kế đơn giản và hiệu quả.

Tuy nhiên, các phương án thực hiện thay thế có thể bao gồm bước lựa chọn hai hoặc nhiều thiết bị âm thanh được trang bị loa và thay đổi quy trình xử lý âm thanh cho ít nhất hai, ít nhất ba (và, trong một số trường hợp, tất cả) trong số hai hoặc nhiều thiết bị âm thanh được trang bị loa. Theo một số ví dụ như vậy, ít nhất một trong số các thay đổi xử lý âm thanh (ví dụ, giảm mức phát lại) đối với thiết bị âm thanh thứ nhất có thể khác với thay đổi xử lý âm thanh đối với thiết bị âm thanh thứ hai. Theo một số ví dụ, sự khác biệt giữa thay đổi xử lý âm thanh có thể, dựa ít nhất một phần vào, vị trí hiện tại được ước lượng của người hoặc vị trí micro so với vị trí của mỗi thiết bị âm thanh. Theo một số phương án thực hiện như vậy, các thay đổi xử lý âm thanh có thể bao gồm bước áp dụng các kích hoạt loa khác nhau tại các vị trí loa khác nhau như một phần của việc thay đổi quy trình kết xuất để làm lệch kết xuất tín hiệu âm thanh khỏi vị trí hiện tại được ước lượng của người quan tâm. Theo một số ví dụ, sự khác biệt giữa các thay đổi xử lý âm thanh có thể dựa ít nhất một phần vào khả năng của loa. Ví dụ, nếu các thay đổi xử lý âm thanh bao gồm bước giảm mức âm thanh trong dải trầm, thì những thay đổi như vậy có thể được áp dụng mạnh mẽ hơn cho thiết bị âm thanh bao gồm một hoặc nhiều loa có khả năng tái tạo âm lượng lớn trong dải trầm.

Tiếp theo, sáng chế sẽ mô tả chi tiết hơn về khía cạnh thay đổi xử lý âm thanh, mà ở đây có thể được gọi là khía cạnh “thay đổi kết xuất”. Phần mô tả này đôi khi có thể đề cập đến khía cạnh này là “giảm âm lượng gần nhất” (ví dụ, giảm âm lượng mà nội dung âm thanh tại đó, sẽ được phát bằng một, hai, ba hoặc nhiều loa gần nhất, được kết xuất), mặc dù (như đã nêu ở phần khác ở đây) nói chung hơn, trong nhiều phương án thực hiện có thể bị ảnh hưởng là một hoặc nhiều thay đổi đối với quy trình xử lý âm

thanh mà được hoặc hướng tới việc cải thiện ước lượng, đo và/hoặc tiêu chí tỷ lệ tín hiệu trên tiếng vọng tổng thể để có khả năng để ghi nhận hoặc cảm nhận bộ phát âm thanh mong muốn (ví dụ, người đã nói từ đánh thức). Trong một số trường hợp, các thay đổi xử lý âm thanh (ví dụ, “giảm” âm lượng của nội dung âm thanh được kết xuất) sẽ hoặc có thể được điều chỉnh bằng tham số của số lượng hiệu ứng liên tục nào đó. Ví dụ, trong ngữ cảnh giảm âm lượng loa, một số phương án thực hiện có thể có khả năng áp dụng lượng suy giảm (dB) có thể điều chỉnh được (ví dụ, có thể điều chỉnh liên tục). Theo một số ví dụ như vậy, lượng suy giảm có thể điều chỉnh được có thể có dải thứ nhất (ví dụ, từ 0 đến 3 dB) đối với thay đổi đủ nhận biết và dải thứ hai (ví dụ, từ 0 đến 20 dB) để cải thiện hiệu quả cụ thể đối với SER nhưng có thể dễ dàng nhận biết đối với người nghe.

Trong một số phương án áp dụng sơ đồ đã nêu (NGŨ CẢNH, QUYẾT ĐỊNH và KẾT XUẤT hoặc THAY ĐỔI KẾT XUẤT), có thể không có ranh giới cứng cụ thể “gần nhất” (ví dụ, đối với loa hoặc thiết bị “gần nhất” với người dùng, hoặc đối với phần tử hệ thống hoặc cá nhân khác), và không làm mất đi tính tổng quát, thay đổi kết xuất có thể là hoặc bao gồm thay đổi (ví dụ, thay đổi liên tục) một hoặc nhiều bước như sau:

A. Chế độ thay đổi đầu ra để giảm đầu ra âm thanh từ một hoặc nhiều thiết bị âm thanh, trong đó (các) thay đổi trong đầu ra âm thanh có thể bao gồm một hoặc nhiều trong số:

i) giảm mức tổng thể của đầu ra thiết bị âm thanh (giảm âm lượng một hoặc nhiều loa, tắt một hoặc nhiều loa);

ii) định hình phổ của đầu ra của một hoặc nhiều loa, ví dụ, bộ lọc cân bằng (equalization – EQ) cơ bản tuyến tính được thiết kế để tạo ra đầu ra khác với phổ của âm thanh mà sáng chế muốn phát hiện. Theo một số ví dụ, nếu phổ đầu ra đang được định hình để phát hiện giọng nói của con người, thì bộ lọc có thể giảm tần số trong dải tần xấp xỉ từ 500 đến 3 kHz (ví dụ, cộng hoặc trừ 5% hoặc 10% ở mỗi đầu của dải tần), hoặc định hình âm lượng để nhấn mạnh các tần số thấp và cao để lại không gian trong các băng tần giữa (ví dụ, trong dải tần xấp xỉ từ 500 đến 3 kHz);

iii) thay đổi các giới hạn trên hoặc đỉnh của đầu ra để giảm mức đỉnh và/hoặc giảm các sản phẩm biến dạng mà có thể làm giảm thêm hiệu suất của

việc khử tiếng vọng bất kỳ là một phần của toàn bộ hệ thống tạo ra SER đạt được để phát hiện âm thanh, ví dụ, bộ nén dải động miền thời gian hoặc bộ nén đa băng tần phụ thuộc tần số. Những sửa đổi tín hiệu âm thanh như vậy có thể giảm một cách hiệu quả biên độ của tín hiệu âm thanh và có thể giúp hạn chế việc lệch của loa;

iv) điều khiển theo không gian âm thanh theo cách có xu hướng giảm năng lượng hoặc kết hợp đầu ra của một hoặc nhiều loa với một hoặc nhiều micrô mà hệ thống (ví dụ, bộ quản lý xử lý âm thanh) kích hoạt SER cao hơn tại đó, ví dụ, như trong các ví dụ “làm lệch” được mô tả ở đây;

v) Sử dụng phương pháp cắt lát thời gian theo thời gian hoặc điều chỉnh để tạo ra ‘các khe’ hoặc các khoảng thời gian đầu ra thời gian tần số thừa thớt hơn đủ để thu được âm thanh thoáng qua, như trong các ví dụ chèn khe được mô tả bên dưới; và/hoặc

vi) Thay đổi âm thanh theo cách kết hợp nào đó của các cách trên; và/hoặc

B. Bảo toàn năng lượng và/hoặc tạo ra tính liên tục tại một tập hợp vị trí lắng nghe rộng hoặc cụ thể, ví dụ, bao gồm một hoặc nhiều thông tin sau:

i) Theo một số ví dụ, năng lượng lấy ra từ một loa có thể được bù lại bằng cách cung cấp năng lượng bổ sung vào hoặc cho loa khác. Trong một số trường hợp, âm lượng tổng thể vẫn giữ nguyên, hoặc về cơ bản là giữ nguyên. Đây không phải là một đặc tính thiết yếu, nhưng có thể là cách hiệu quả cho phép thay đổi mạnh hơn đối với việc xử lý âm thanh của ‘thiết bị gần nhất’ hoặc tập hợp thiết bị gần nhất’ mà không làm mất nội dung. Tuy nhiên, tính liên tục và/hoặc bảo toàn năng lượng có thể đặc biệt thích hợp khi xử lý các cảnh âm thanh và đầu ra âm thanh phức tạp; và/hoặc

ii) Hằng số thời gian của quy trình kích hoạt, cụ thể là các thay đổi đối với xử lý âm thanh có thể được áp dụng nhanh hơn một chút (ví dụ, từ 100 đến 200 ms) so với khi chúng được đưa về trạng thái bình thường (ví dụ, từ 1000 đến 10000 ms) sao cho (các) thay đổi trong quá trình xử lý âm thanh, nếu nhận biết được, dường như có chủ ý, nhưng sự trở lại sau đó từ (các) thay đổi có thể dường như không liên quan đến bất kỳ sự kiện hoặc thay đổi thực tế nào (từ quan điểm

của người dùng) và, trong một số trường hợp, có thể chậm đến mức khó nhận biết được.

Các ví dụ khác về cách thức ngữ cảnh và quyết định có thể được trình bày.

Phương án A:

(NGŨ CẢNH) Theo một ví dụ, thông tin ngữ cảnh có thể được lập công thức toán học như sau:

$H(a, b)$, khoảng cách vật lý gần đúng tính bằng mét giữa các thiết bị **a** và **b**:

$$H(a, b) \neq 0 \in \mathbb{R}$$

$$\forall a, b \in D$$

trong đó D đại diện cho tập hợp tất cả các thiết bị trong hệ thống. S , SER được ước lượng tại mỗi thiết bị, có thể được biểu thị như sau:

$$S(a) \in \mathbb{R} \quad \forall a \in D$$

Xác định H & S:

H là thuộc tính của vị trí vật lý của các thiết bị, và do đó có thể được xác định hoặc được ước lượng bằng:

(1) Chỉ báo trực tiếp bởi người dùng, ví dụ, bằng cách sử dụng điện thoại thông minh hoặc thiết bị máy tính bảng để đánh dấu hoặc biểu thị vị trí gần đúng của các thiết bị trên sơ đồ mặt bằng hoặc dạng biểu diễn sơ đồ tương tự của môi trường. Các giao diện số như vậy đã trở nên phổ biến trong việc quản lý cấu hình, nhóm, tên, mục đích, và định danh của các thiết bị nhà thông minh. Ví dụ, chỉ báo trực tiếp như vậy có thể được cung cấp qua ứng dụng điện thoại thông minh Amazon Alexa, ứng dụng bộ điều khiển Sonos S2, hoặc ứng dụng tương tự.

(2) Giải bài toán phép đo tam giác cơ bản bằng cách sử dụng cường độ tín hiệu đo được (đôi khi được gọi là Chỉ báo cường độ tín hiệu nhận được hay RSSI) của các công nghệ truyền thông không dây phổ biến như Bluetooth, Wi-Fi, ZigBee, v.v., để đưa ra các ước lượng về khoảng cách vật lý giữa thiết bị, ví dụ, như được mô tả trong tài liệu của J. Yang và Y. Chen, “Indoor Localization Using Improved RSS-Based Lateration Methods”, *GLOBECOM 2009 – 2009 Hội nghị viễn thông toàn cầu IEEE, Honolulu, HI*, 2009, trang 1 đến 6, DOI: 10.1109/GLOCOM.2009.5425237 và/hoặc được mô tả trong tài liệu của Mardeni, R. & Othman, Shaifull & Nizam, (2010) “Node Positioning

in ZigBee Network Using Trilateration Method Based on the Received Signal Strength Indicator (RSSI)” 46, cả hai tài liệu này đều được đưa vào đây bằng cách viện dẫn.

$S(a)$ là ước lượng của tỷ lệ lời nói trên tiếng vọng tại thiết bị a . Theo định nghĩa, tỷ lệ lời nói trên tiếng vọng tính bằng dB được xác định bằng:

$$S(a) = \widehat{\mu_{sp}} - \widehat{\mu_E}$$

Trong phương trình trên, $\widehat{\mu_{sp}}$ biểu diễn ước lượng năng lượng lời nói tính bằng dB, và $\widehat{\mu_E}$ biểu diễn ước lượng năng lượng tiếng vọng dư sau khi khử tiếng vọng, tính bằng dB. Các phương pháp luận khác nhau để ước lượng các đại lượng này được mô tả ở đây, ví dụ:

Năng lượng lời nói và năng lượng tiếng vọng dư có thể được ước lượng bằng quy trình đo ngoại tuyến được thực hiện cho một thiết bị cụ thể, có tính đến sự gần âm thanh giữa micrô và loa của thiết bị, cũng như hiệu suất của mạch khử tiếng vọng trên bo mạch. Theo một số ví dụ như vậy, mức năng lượng lời nói trung bình “AvgSpeech” có thể được xác định bằng mức năng lượng lời nói trung bình của con người khi được đo bằng thiết bị ở khoảng cách danh định. Ví dụ, lời nói của một số ít người đứng cách thiết bị được trang bị micrô 1m có thể được thiết bị ghi lại trong quá trình sản xuất và năng lượng có thể được tính trung bình để tạo ra AvgSpeech. Theo một số ví dụ như vậy, mức năng lượng tiếng vọng dư trung bình “AvgEcho” có thể được ước lượng bằng cách phát nội dung âm nhạc từ thiết bị trong quá trình sản xuất và chạy mạch khử tiếng vọng trên bo mạch để tạo ra tín hiệu dư tiếng vọng. Có thể sử dụng phép lấy trung bình năng lượng của tín hiệu dư tiếng vọng cho mẫu nội dung âm nhạc nhỏ để ước lượng AvgEcho. Khi thiết bị không phát âm thanh, thay vào đó, AvgEcho có thể được thiết lập thành giá trị thấp danh định, như -96,0 dB. Theo một số phương án thực hiện như vậy, năng lượng lời nói và năng lượng tiếng vọng dư có thể được biểu diễn như sau:

$$\widehat{\mu_{sp}} = AvgSpeech_{dB}$$

$$\widehat{\mu_E} = AvgEcho_{dB}$$

(2) Theo một số ví dụ, năng lượng lời nói trung bình có thể được xác định bằng cách lấy năng lượng của tín hiệu micrô tương ứng với tiếng nói của người dùng được xác định bằng bộ phát hiện hoạt động giọng nói (voice-activity-detector – VAD). Theo một số ví dụ như vậy, năng lượng tiếng vọng dư trung bình có thể được ước lượng bởi năng lượng của tín hiệu micrô khi VAD không biểu thị lời nói. Nếu x biểu diễn các mẫu

điều chế mã xung (pulse-code modulation – PCM) micrô của thiết bị a ở tốc độ lấy mẫu nào đó, và V biểu diễn cờ VAD lấy giá trị 1,0 cho các mẫu tương ứng với hoạt động giọng nói, và nếu không thì bằng 0,0, năng lượng lời nói và năng lượng tiếng vọng dư có thể được biểu diễn như sau:

$$\widehat{\mu_{Sp}} = 10 * \log_{10} \left(\sum_n V[n] x[n]^2 \right)$$

$$\widehat{\mu_E} = 10 * \log_{10} \left(\sum_n (1 - V[n]) x[n]^2 \right)$$

(3) Ngoài các phương pháp trước đó, theo một số phương án thực hiện, năng lượng trong micrô có thể được xem là biến ngẫu nhiên và được lập mô hình riêng dựa trên phép xác định VAD. Các mô hình thống kê Sp và E của năng lượng lời nói và tiếng vọng có thể được ước lượng lần lượt bằng cách sử dụng số lượng kỹ thuật mô hình thống kê bất kỳ. Giá trị trung bình tính bằng dB đối với cả lời nói và tiếng vọng để tính xấp xỉ $S(a)$ sau đó có thể được rút ra lần lượt từ Sp và E . Các phương pháp chung để thu được quy trình này được tìm thấy trong lĩnh vực xử lý tín hiệu thống kê, ví dụ:

- Giả sử phân phối Gauss của năng lượng và tính toán thống kê bậc hai sai lệch $\hat{\mu}$ và $\hat{\sigma}$.
- Xây dựng biểu đồ giá trị năng lượng được phân loại riêng để đạt được việc phân phối đa chế độ có thể có, sau khi áp dụng bước ước lượng tham số kỳ vọng–tối đa hóa (expectation–maximization – EM) cho mô hình hỗn hợp (ví dụ, mô hình hỗn hợp Gauss), giá trị trung bình lớn nhất $\hat{\mu}$ thuộc về bất kỳ trong số các phân bố con trong hỗn hợp có thể được dùng.

(QUYẾT ĐỊNH) Như đã lưu ý ở phần khác trong bản mô tả này, trong một số phương án thực hiện được mô tả khác nhau, khía cạnh quyết định xác định thiết bị nào nhận được sửa đổi xử lý âm thanh, như sửa đổi kết xuất, và theo một số phương án là chỉ báo về mức độ sửa đổi cần thiết để cải thiện SER mong muốn. Một số phương án như vậy có thể được tạo cấu hình để cải thiện SER tại thiết bị có giá trị SER ban đầu tốt nhất, ví dụ, như được xác định bằng cách tìm giá trị S lớn nhất trên tất cả các thiết bị trong tập hợp D. Các phương án khác có thể được tạo cấu hình để cải thiện SER khi có cơ hội tại thiết bị được người dùng thường xuyên sử dụng, được xác định dựa trên các mẫu sử dụng trước đây. Các phương án khác có thể được tạo cấu hình để cố gắng cải

thiện SER tại nhiều vị trí micrô, ví dụ, lựa chọn nhiều thiết bị cho các mục đích của việc mô tả dưới đây.

Khi một hoặc nhiều vị trí micrô được xác định, theo một số phương án thực hiện như vậy, quy trình cải tiến SER (SER improvement – SERI) mong muốn có thể được xác định như sau:

$$SERI = S(m) - TargetSER [dB]$$

Trong phương trình trên, m biểu diễn vị trí thiết bị/micrô đang được cải thiện, và TargetSER biểu diễn ngưỡng, có thể được thiết lập bằng ứng dụng đang sử dụng. Ví dụ, thuật toán phát hiện từ đánh thức có thể cho phép SER hoạt động thấp hơn so với bộ nhận dạng lời nói có từ vựng lớn. Các giá trị điển hình cho TargetSER có thể trong khoảng từ -6 dB đến 12 dB. Như đã nêu, theo một số phương án, nếu S(m) không được biết hoặc được ước lượng dễ dàng, thì giá trị thiết lập trước có thể đủ dựa trên các số đo ngoại tuyến của lời nói và tiếng vọng được ghi lại trong phòng hoặc thiết lập tiếng vọng điển hình. Một số phương án có thể xác định các thiết bị trong đó việc xử lý âm thanh (ví dụ, kết xuất) sẽ được sửa đổi bằng cách xác định f_n nằm trong khoảng từ 0 đến 1. Các phương án khác có thể bao gồm bước xác định mức mà việc xử lý âm thanh (ví dụ, kết xuất) cần được sửa đổi theo đơn vị decibels cải thiện tỷ lệ lời nói trên tiếng vọng, s_n , có khả năng được tính toán theo:

$$s_n = SERI * f_n$$

Một số phương án có thể tính f_n trực tiếp từ hình dạng thiết bị, ví dụ, như sau:

$$f_n = 1 - \left(\frac{H(m, n)}{\max_{i \in D} H(m, i)} \right)^2$$

Trong phương trình trên, m biểu diễn chỉ số của thiết bị sẽ được lựa chọn để sửa đổi xử lý âm thanh (ví dụ, kết xuất) lớn nhất, như lưu ý ở trên. Các phương án thực hiện khác có thể bao gồm các lựa chọn khác về các hàm nói lỏng hoặc làm mịn trên hình dạng thiết bị.

Phương án B (tham chiếu đến vùng người dùng):

Theo một số phương án, bối cảnh và các khía cạnh quyết định của sáng chế sẽ được thực hiện trong bối cảnh của một hoặc nhiều vùng người dùng. Như được mô tả chi tiết ở phần sau của sáng chế này, tập hợp các đặc điểm âm học $\mathbf{W}(j) = [\mathbf{w}_1^T(j) \dots \mathbf{w}_N^T(j)]^T$ có thể được dùng để ước lượng xác suất hậu nghiệm $p(C_k | \mathbf{W}(j))$ cho

một số tập hợp nhân vùng C_k , đối với $k = \{1 \dots K\}$, cho K vùng người dùng khác nhau trong môi trường. Sự kết hợp của từng thiết bị âm thanh với từng vùng người dùng có thể được cung cấp bởi chính người dùng như một phần của quy trình huấn luyện được mô tả trong sáng chế này, hoặc theo cách khác thông qua phương tiện của ứng dụng, ví dụ, ứng dụng điện thoại thông minh Alexa hoặc ứng dụng điện thoại thông minh bộ điều khiển Sonos S2. Ví dụ, một số phương án thực hiện có thể biểu thị việc kết hợp của thiết bị thứ j với vùng người dùng có nhân vùng C_k là $z(C_k, n) \in [0, 1]$. Theo một số phương án, cả $z(C_k, n)$ và các xác suất hậu nghiệm $p(C_k | \mathbf{W}(j))$ có thể được xem như là thông tin ngữ cảnh. Thay vào đó, một số phương án có thể coi chính các đặc tính âm thanh $\mathbf{W}(j)$ là một phần của ngữ cảnh. Theo các phương án khác, nhiều hơn một trong các đại lượng ($z(C_k, n)$, các xác suất hậu nghiệm $p(C_k | \mathbf{W}(j))$ và chính các đặc tính âm thanh $\mathbf{W}(j)$) và/hoặc sự kết hợp của các đại lượng này có thể là một phần của thông tin ngữ cảnh.

Khía cạnh quyết định của các phương án khác nhau có thể sử dụng các đại lượng liên quan đến một hoặc nhiều vùng người dùng trong lựa chọn thiết bị. Nếu cả z và p đều có sẵn, thì quyết định làm ví dụ có thể được đưa ra là:

$$f_n = \max_k (z(C_k, n) p(C_k | \mathbf{W}(j)))^\delta \quad k = \{1 \dots K\}$$

Theo các phương án như vậy, các thiết bị có liên kết cao nhất với vùng người dùng có nhiều khả năng chứa người dùng nhất sẽ có nhiều thay đổi xử lý âm thanh (ví dụ, kết xuất) được áp dụng cho chúng. Theo một số ví dụ, δ có thể là số dương trong phạm vi $[0, 5, 4, 0]$. Theo một số ví dụ như vậy, δ có thể được dùng để kiểm soát phạm vi của sự thay đổi kết xuất trong không gian. Trong các phương án thực hiện như vậy, nếu δ được lựa chọn là 0,5, thì nhiều thiết bị sẽ nhận được thay đổi kết xuất lớn hơn, còn giá trị 4,0 sẽ hạn chế thay đổi kết xuất chỉ với các thiết bị gần với vùng người dùng có khả năng nhất.

Các tác giả sáng chế cũng dự kiến một loại phương án khác, trong đó các đặc tính âm thanh $\mathbf{W}(j)$ được dùng trực tiếp trong khía cạnh quyết định. Ví dụ, nếu độ tin cậy của từ đánh thức được kết hợp với tiếng nói j là $w_n(j)$, thì bước lựa chọn thiết bị có thể được thực hiện theo phương trình sau:

$$f_n = \left(\frac{w_n(j)}{\max_k w_k(j)} \right)^\delta \quad k = \{1 \dots K\}$$

Trong các phương trình trên, δ có cách giải thích giống như ví dụ trước, và còn có tác dụng bù cho phân phối điển hình của có độ tin cậy của từ đánh thức mà có thể tăng lên đối với hệ thống từ đánh thức cụ thể. Nếu hầu hết các thiết bị có xu hướng báo cáo độ tin cậy từ đánh thức cao, δ có thể được lựa chọn là số tương đối cao hơn, như 3,0, để tăng tính đặc trưng về không gian của ứng dụng thay đổi kết xuất. Nếu độ tin cậy từ đánh thức có xu hướng giảm nhanh khi người dùng ở xa thiết bị hơn, δ có thể được lựa chọn là số tương đối thấp hơn, như 1,0 hoặc thậm chí 0,5 để bao gồm nhiều thiết bị hơn vào ứng dụng thay đổi kết xuất. Người có hiểu biết trung bình trong lĩnh vực kỹ thuật này sẽ đánh giá cao rằng theo một số phương án thực hiện thay thế, công thức tương tự như một trong số các đặc tính âm thanh ở trên, như ước lượng mức lời nói ở micro của thiết bị, và/hoặc tỷ lệ âm trực tiếp trên âm dội của tiếng nói của người dùng có thể được thay thế cho độ tin cậy của từ đánh thức.

Fig.2A là sơ đồ khối thể hiện các ví dụ về các thành phần của thiết bị hoặc hệ thống có khả năng thực hiện các khía cạnh khác nhau của sáng chế. Cũng như các hình vẽ khác được cung cấp ở đây, các loại và số lượng các phần tử thể hiện trên Fig.2A chỉ được cung cấp theo cách ví dụ. Các phương án thực hiện khác có thể bao gồm nhiều hơn, ít hơn và/hoặc các loại và số lượng phần tử khác nhau. Theo một số ví dụ, thiết bị 200 có thể là, hoặc có thể bao gồm, thiết bị được tạo cấu hình để thực hiện ít nhất một số trong số các phương pháp được mô tả ở đây. Theo một số phương án thực hiện, thiết bị 200 có thể là, hoặc có thể bao gồm, loa thông minh, máy tính xách tay, điện thoại di động, thiết bị máy tính bảng, trung tâm nhà thông minh hoặc thiết bị khác được tạo cấu hình để thực hiện ít nhất một số trong số các phương pháp được mô tả ở đây. Theo một số phương án thực hiện, thiết bị 200 có thể được tạo cấu hình để áp dụng bộ quản lý phiên âm thanh. Theo một số phương án thực hiện như vậy, thiết bị 200 có thể là, hoặc có thể bao gồm, máy chủ.

Trong ví dụ này, thiết bị 200 bao gồm hệ thống giao diện 205 và hệ thống điều khiển 210. Hệ thống giao diện 205 có thể, theo một số phương án thực hiện, được tạo cấu hình để truyền thông với một hoặc nhiều thiết bị đang thực thi, hoặc được tạo cấu hình để thực thi, các ứng dụng phần mềm. Các ứng dụng phần mềm như vậy đôi khi có

thể được gọi ở đây là các “ứng dụng” hoặc đơn giản là các “app”. Theo một số phương án thực hiện, hệ thống giao diện 205, có thể được tạo cấu hình để trao đổi thông tin điều khiển và dữ liệu liên quan với các ứng dụng. Theo một số phương án thực hiện, hệ thống giao diện 205, có thể được tạo cấu hình để truyền thông với một hoặc nhiều thiết bị khác của môi trường âm thanh. Theo một số ví dụ, môi trường âm thanh có thể là môi trường âm thanh gia đình. Theo một số phương án thực hiện, hệ thống giao diện 205 có thể được tạo cấu hình để trao đổi thông tin điều khiển và dữ liệu liên quan với các thiết bị âm thanh của môi trường âm thanh. Theo một số ví dụ, thông tin điều khiển và dữ liệu liên quan, có thể liên quan đến một hoặc nhiều ứng dụng mà thiết bị 200 được tạo cấu hình để truyền thông.

Theo một số phương án thực hiện, hệ thống giao diện 205 có thể được tạo cấu hình để nhận dữ liệu âm thanh. Dữ liệu âm thanh có thể bao gồm các tín hiệu âm thanh được lập lịch để tái tạo bằng ít nhất một số loa của môi trường âm thanh. Dữ liệu âm thanh có thể bao gồm một hoặc nhiều tín hiệu âm thanh và dữ liệu không gian liên quan. Ví dụ, dữ liệu không gian có thể bao gồm dữ liệu kênh và/hoặc siêu dữ liệu không gian. Hệ thống giao diện 205 có thể được tạo cấu hình để cung cấp tín hiệu âm thanh được kết xuất cho ít nhất một số loa trong tập hợp loa của môi trường. Theo một số phương án thực hiện, hệ thống giao diện 205 có thể được tạo cấu hình để nhận đầu vào từ một hoặc nhiều micro trong môi trường.

Hệ thống giao diện 205 có thể bao gồm một hoặc nhiều giao diện mạng và/hoặc một hoặc nhiều giao diện thiết bị bên ngoài (như một hoặc nhiều giao diện bus nối tiếp đa năng (universal serial bus – USB)). Theo một số phương án thực hiện, hệ thống giao diện 205 có thể bao gồm một hoặc nhiều giao diện không dây. Hệ thống giao diện 205 có thể bao gồm một hoặc nhiều thiết bị áp dụng giao diện người dùng, như một hoặc nhiều micro, một hoặc nhiều loa, hệ thống hiển thị, hệ thống cảm biến chạm và/hoặc hệ thống cảm biến cử chỉ. Theo một số ví dụ, hệ thống giao diện 205 có thể bao gồm một hoặc nhiều giao diện giữa hệ thống điều khiển 210 và hệ thống bộ nhớ, như hệ thống bộ nhớ tùy chọn 215 được thể hiện trên Fig.2A. Tuy nhiên, hệ thống điều khiển 210 có thể bao gồm hệ thống bộ nhớ trong một số trường hợp.

Ví dụ, hệ thống điều khiển 210 có thể bao gồm bộ xử lý một chip hoặc nhiều chip đa năng, bộ xử lý tín hiệu số (digital signal processor – DSP), mạch tích hợp chuyên

dụng (application specific integrated circuit – ASIC), mảng cổng lập trình được theo trường (field programmable gate array – FPGA) hoặc thiết bị logic có thể lập trình được khác, cổng rời rạc hoặc logic bán dẫn, và/hoặc các thành phần phần cứng rời rạc.

Theo một số phương án thực hiện, hệ thống điều khiển 210 có thể nằm trong nhiều hơn một thiết bị. Ví dụ, một phần của hệ thống điều khiển 210 có thể nằm trong thiết bị trong một trong số các môi trường được mô tả ở đây và phần khác của hệ thống điều khiển 210 có thể nằm trong thiết bị bên ngoài môi trường, như máy chủ, thiết bị di động (ví dụ, điện thoại thông minh hoặc máy tính bảng), v.v. Theo các ví dụ khác, một phần của hệ thống điều khiển 210 có thể nằm trong một thiết bị trong một trong số các môi trường được mô tả ở đây và phần khác của hệ thống điều khiển 210 có thể nằm trong một hoặc nhiều thiết bị khác của môi trường. Ví dụ, chức năng hệ thống điều khiển có thể được phân phối trên nhiều thiết bị âm thanh thông minh của một môi trường, hoặc có thể được thiết bị hòa âm dùng chung (như có thể được gọi ở đây là bộ quản lý phiên âm thanh hoặc trung tâm nhà thông minh) và một hoặc nhiều thiết bị khác của môi trường. Theo một số ví dụ như vậy, hệ thống giao diện 205 cũng có thể nằm trong nhiều hơn một thiết bị.

Theo một số phương án thực hiện, hệ thống điều khiển 210 có thể được tạo cấu hình để thực hiện, ít nhất một phần, phương pháp được mô tả ở đây. Theo một số ví dụ, hệ thống điều khiển 210 có thể được tạo cấu hình để áp dụng các phương pháp quản lý phiên âm thanh, mà trong một số trường hợp có thể bao gồm bước xác định một hoặc nhiều kiểu thay đổi xử lý âm thanh để áp dụng cho dữ liệu âm thanh được kết xuất thành tín hiệu cấp cho loa cho hai hoặc nhiều âm thanh thiết bị của môi trường âm thanh. Theo một số phương án thực hiện, các thay đổi xử lý âm thanh có thể có tác dụng tăng tỷ lệ lời nói trên tiếng vọng ở một hoặc nhiều micrô trong môi trường âm thanh.

Một số hoặc tất cả các phương pháp được mô tả ở đây có thể được thực hiện bằng một hoặc nhiều thiết bị theo các lệnh (ví dụ, phần mềm) được lưu trữ trên một hoặc nhiều phương tiện bất biến. Phương tiện bất biến như vậy có thể bao gồm các thiết bị nhớ như được mô tả ở đây, bao gồm nhưng không giới hạn ở các thiết bị nhớ truy cập ngẫu nhiên (RAM), thiết bị nhớ chỉ đọc (ROM), v.v. Một hoặc nhiều phương tiện bất biến có thể, ví dụ, nằm trong hệ thống nhớ tùy chọn 215 được thể hiện trên Fig.2A và/hoặc trong hệ thống điều khiển 210. Do đó, các khía cạnh đối mới khác nhau của các

đối tượng được mô tả trong sáng chế này có thể được thực hiện trong một hoặc nhiều phương tiện bất biến có phần mềm được lưu trữ trên đó. Ví dụ, phần mềm có thể bao gồm các các lệnh để điều khiển ít nhất một thiết bị để áp dụng các phương pháp quản lý phiên âm thanh. Theo một số ví dụ, phần mềm có thể bao gồm các lệnh để điều khiển một hoặc nhiều thiết bị âm thanh của môi trường âm thanh để thu được, xử lý và/hoặc cung cấp dữ liệu âm thanh. Theo một số ví dụ, phần mềm có thể bao gồm các các lệnh để xác định một hoặc nhiều kiểu thay đổi xử lý âm thanh để áp dụng cho dữ liệu âm thanh được kết xuất thành tín hiệu cấp cho loa cho hai hoặc nhiều thiết bị âm thanh của môi trường âm thanh. Theo một số phương án thực hiện, các thay đổi xử lý âm thanh có thể có tác dụng tăng tỷ lệ lời nói trên tiếng vọng ở một hoặc nhiều micrô trong môi trường âm thanh. Ví dụ, phần mềm có thể thực thi được bởi một hoặc nhiều thành phần của hệ thống điều khiển, như hệ thống điều khiển 210 trên Fig.2A.

Theo một số ví dụ, thiết bị 200 có thể bao gồm hệ thống micrô tùy chọn 220 thể hiện trên Fig.2A. Hệ thống micrô tùy chọn 220 có thể bao gồm một hoặc nhiều micrô. Theo một số phương án thực hiện, một hoặc nhiều micrô có thể là một phần của, hoặc được kết hợp với, thiết bị khác, như loa của hệ thống loa, thiết bị âm thanh thông minh, v.v. Theo một số ví dụ, thiết bị 200 có thể không bao gồm hệ thống micrô 220. Tuy nhiên, theo một số phương án thực hiện như vậy, thiết bị 200 vẫn có thể được tạo cấu hình để nhận dữ liệu micrô cho một hoặc nhiều micrô trong môi trường âm thanh qua hệ thống giao diện 210.

Theo một số phương án thực hiện, thiết bị 200 có thể bao gồm hệ thống loa tùy chọn 225 được thể hiện trên Fig.2A. Hệ thống loa tùy chọn 225 có thể bao gồm một hoặc nhiều loa. Đôi khi loa có thể được gọi ở đây là “loa phóng thanh”. Theo một số ví dụ, ít nhất một số loa của hệ thống loa tùy chọn 225 có thể được bố trí tùy ý. Ví dụ, ít nhất một số loa của hệ thống loa tùy chọn 225 có thể được bố trí ở các vị trí không tương ứng với bất kỳ sơ đồ loa tiêu chuẩn nào được quy định, như Dolby 5.1, Dolby 5.1.2, Dolby 7.1, Dolby 7.1.4, Dolby 9.1, Hamasaki 22.2, v.v.. Theo một số ví dụ như vậy, ít nhất một số loa của hệ thống loa tùy chọn 225 có thể được đặt ở những vị trí thuận tiện cho không gian (ví dụ, ở những vị trí có không gian để chứa các loa), nhưng không phải là bất kỳ sơ đồ loa tiêu chuẩn quy định nào. Theo một số ví dụ, thiết bị 200 có thể không bao gồm hệ thống loa tùy chọn 225.

Theo một số phương án thực hiện, thiết bị 200 có thể bao gồm hệ thống cảm biến tùy chọn 230 được thể hiện trên Fig.2A. Hệ thống cảm biến tùy chọn 230 có thể bao gồm một hoặc nhiều camera, cảm biến chạm, cảm biến cử chỉ, cảm biến chuyển động, v.v. Theo một số phương án thực hiện, hệ thống cảm biến tùy chọn 230 có thể bao gồm một hoặc nhiều camera. Theo một số phương án thực hiện, camera có thể là camera không có giá đỡ. Theo một số ví dụ, một hoặc nhiều camera của hệ thống cảm biến tùy chọn 230 có thể nằm trong thiết bị âm thanh thông minh mà có thể là thiết bị âm thanh chuyên dùng hoặc trợ lý ảo. Theo một số ví dụ như vậy, một hoặc nhiều camera của hệ thống cảm biến tùy chọn 230 có thể nằm trong TV, điện thoại di động hoặc loa thông minh. Theo một số ví dụ, thiết bị 200 có thể không bao gồm hệ thống cảm biến 230. Tuy nhiên, theo một số phương án thực hiện như vậy, thiết bị 200 vẫn có thể được tạo cấu hình để nhận dữ liệu cảm biến cho một hoặc nhiều cảm biến trong môi trường âm thanh qua hệ thống giao diện 210.

Theo một số phương án thực hiện, thiết bị 200 có thể bao gồm hệ thống hiển thị tùy ý 235 thể hiện trên Fig.2A. Hệ thống hiển thị tùy ý 235 có thể bao gồm một hoặc nhiều màn hình, như một hoặc nhiều màn hình đi-ốt phát quang (light-emitting diode – LED). Trong một số trường hợp, hệ thống hiển thị tùy ý 235 có thể bao gồm một hoặc nhiều màn hình đi-ốt phát quang hữu cơ (organic light-emitting diode – OLED). Theo một số ví dụ, trong đó thiết bị 200 bao gồm hệ thống hiển thị 235, hệ thống cảm biến 230 có thể bao gồm hệ thống cảm biến chạm và/hoặc hệ thống cảm biến cử chỉ đặt gần một hoặc nhiều màn hình của hệ thống hiển thị 235. Theo một số phương án thực hiện như vậy, hệ thống điều khiển 210 có thể được tạo cấu hình để điều khiển hệ thống hiển thị 235 để trình diễn một hoặc nhiều giao diện người dùng đồ họa (graphical user interface – GUI).

Theo một số ví dụ, thiết bị 200 có thể là, hoặc có thể bao gồm, thiết bị âm thanh thông minh. Theo một số phương án thực hiện như vậy, thiết bị 200 có thể là, hoặc có thể áp dụng (ít nhất một phần), bộ phát hiện từ đánh thức. Ví dụ, thiết bị 200 có thể là, hoặc có thể áp dụng (ít nhất một phần), trợ lý ảo.

Fig.2B là lưu đồ bao gồm các khối của phương pháp quản lý phiên âm thanh theo một số phương án thực hiện. Các khối của phương pháp 250, giống như các phương pháp khác được mô tả ở đây, không nhất thiết phải được thực hiện theo thứ tự được biểu

thị. Theo một số phương án thực hiện, một hoặc nhiều khối của phương pháp 250 có thể được thực hiện đồng thời. Hơn nữa, một số phương án thực hiện của phương pháp 250 có thể bao gồm nhiều hoặc ít khối hơn so với được thể hiện và/hoặc mô tả. Các khối của phương pháp 250 có thể được thực hiện bằng một hoặc nhiều thiết bị, có thể là (hoặc có thể bao gồm) hệ thống điều khiển như hệ thống điều khiển 210 được thể hiện trên Fig.2A và được mô tả ở trên, hoặc một trong các ví dụ về hệ thống điều khiển được mô tả khác. Theo một số phương án thực hiện, các khối của phương pháp 250 có thể được thực hiện, ít nhất một phần, bằng thiết bị áp dụng mà ở đây có thể được gọi là bộ quản lý phiên âm thanh.

Theo ví dụ này, khối 255 bao gồm bước nhận tín hiệu đầu ra từ mỗi micrô trong số nhiều micrô trong môi trường âm thanh. Theo ví dụ này, mỗi micrô trong số nhiều micrô nằm ở vị trí micrô của môi trường âm thanh và tín hiệu đầu ra bao gồm các tín hiệu tương ứng với tiếng nói hiện tại của người. Trong một số trường hợp, tiếng nói hiện tại có thể là tiếng nói từ đánh thức. Tuy nhiên, các tín hiệu đầu ra cũng có thể bao gồm tín hiệu tương ứng với các thời điểm mà trong đó không có người nói. Ví dụ, các tín hiệu như vậy, có thể được dùng để thiết lập các mức cơ bản của tiếng vọng, tiếng ồn, v.v.

Trong ví dụ này, khối 260 bao gồm bước xác định, dựa trên các tín hiệu đầu ra, một hoặc nhiều khía cạnh của thông tin ngữ cảnh liên quan đến con người. Theo phương án thực hiện này, thông tin ngữ cảnh bao gồm vị trí hiện tại được ước lượng của người đó và/hoặc độ gần hiện tại được ước lượng của người đó với một hoặc nhiều vị trí micrô. Như lưu ý ở trên, cụm từ “vị trí micrô” như được dùng ở đây biểu thị vị trí của một hoặc nhiều micrô. Theo một số ví dụ, vị trí của một micrô có thể tương ứng với mảng micrô nằm trong một thiết bị âm thanh. Ví dụ, vị trí micrô có thể là vị trí duy nhất tương ứng với toàn bộ thiết bị âm thanh bao gồm một hoặc nhiều micrô. Theo một số ví dụ như vậy, vị trí micrô có thể là một vị trí tương ứng với trung tâm của mảng micrô của thiết bị âm thanh. Tuy nhiên, trong một số trường hợp, vị trí micrô có thể là vị trí của một micrô duy nhất. Theo một số ví dụ như vậy, thiết bị âm thanh có thể chỉ có một micrô.

Theo một số ví dụ, bước xác định thông tin ngữ cảnh có thể bao gồm bước tạo ra ước lượng về vùng người dùng mà người đó hiện đang ở. Một số ví dụ như vậy có thể bao gồm bước xác định nhiều đặc tính âm thanh hiện tại từ tín hiệu đầu ra của mỗi micrô

và áp dụng bộ phân loại cho nhiều đặc tính âm thanh hiện tại. Ví dụ, bước áp dụng bộ phân loại có thể bao gồm bước áp dụng mô hình được huấn luyện trên các đặc tính âm thanh được xác định từ trước thu được từ nhiều tiếng nói trước đó của con người trong nhiều vùng người dùng trong môi trường. Theo một số ví dụ như vậy, bước xác định một hoặc nhiều khía cạnh thông tin ngữ cảnh liên quan đến con người có thể bao gồm bước xác định, dựa ít nhất một phần vào đầu ra từ bộ phân loại, ước lượng về vùng người dùng mà người đó hiện đang ở. Theo một số ví dụ như vậy, ước lượng về vùng người dùng có thể được xác định mà không cần tham chiếu đến vị trí hình học của nhiều micro. Theo một số ví dụ, tiếng nói hiện tại và những tiếng nói trước đó có thể là, hoặc có thể bao gồm tiếng nói từ đánh thức.

Theo phương án thực hiện này, khối 265 bao gồm bước lựa chọn hai hoặc nhiều thiết bị âm thanh của môi trường âm thanh dựa, ít nhất một phần, vào một hoặc nhiều khía cạnh của thông tin ngữ cảnh, mỗi trong số hai hoặc nhiều thiết bị âm thanh bao gồm ít nhất một loa. Theo một số ví dụ, bước lựa chọn hai hoặc nhiều thiết bị âm thanh của môi trường âm thanh có thể bao gồm bước lựa chọn N thiết bị âm thanh được trang bị loa của môi trường âm thanh, N là số nguyên lớn hơn 2. Trong một số trường hợp, bước lựa chọn hai hoặc nhiều thiết bị âm thanh của môi trường âm thanh, hoặc bước lựa chọn N thiết bị âm thanh được trang bị loa của môi trường âm thanh, có thể bao gồm bước lựa chọn tất cả các thiết bị âm thanh được trang bị loa của môi trường âm thanh.

Theo một số ví dụ, bước lựa chọn hai hoặc nhiều thiết bị âm thanh của môi trường âm thanh có thể được dựa, ít nhất một phần, vào vị trí hiện tại được ước lượng của người so với vị trí micro và/hoặc vị trí thiết bị âm thanh được trang bị loa. Một số ví dụ như vậy có thể bao gồm bước xác định thiết bị âm thanh được trang bị loa gần nhất với vị trí hiện tại được ước lượng của người đó hoặc với vị trí micro gần nhất với vị trí hiện tại được ước lượng của người đó. Theo một số ví dụ như vậy, hai hoặc nhiều thiết bị âm thanh có thể bao gồm thiết bị âm thanh được trang bị loa gần nhất.

Theo một số phương án thực hiện, bước lựa chọn hai hoặc nhiều thiết bị âm thanh có thể được dựa, ít nhất một phần, vào bước xác định xem ước lượng tỷ lệ tín hiệu trên tiếng vọng có nhỏ hơn hoặc bằng ngưỡng tỷ lệ tín hiệu trên tiếng vọng hay không.

Theo ví dụ này, khối 270 bao gồm bước xác định một hoặc nhiều kiểu thay đổi xử lý âm thanh để áp dụng cho dữ liệu âm thanh được kết xuất thành tín hiệu cấp cho

loa cho hai hoặc nhiều thiết bị âm thanh. Theo phương án thực hiện này, các thay đổi xử lý âm thanh có tác dụng tăng tỷ lệ lời nói trên tiếng vọng ở một hoặc nhiều micrô. Theo một số ví dụ, một hoặc nhiều micrô có thể nằm trong nhiều thiết bị âm thanh của môi trường âm thanh. Tuy nhiên, theo một số phương án thực hiện, một hoặc nhiều micrô có thể nằm trong một thiết bị âm thanh của môi trường âm thanh. Theo một số ví dụ, (các) thay đổi xử lý âm thanh có thể khiến giảm mức tái tạo loa cho các loa của hai hoặc nhiều thiết bị âm thanh.

Theo một số ví dụ, ít nhất một trong số các thay đổi xử lý âm thanh đối với thiết bị âm thanh thứ nhất có thể khác với các thay đổi xử lý âm thanh đối với thiết bị âm thanh thứ hai. Ví dụ, (các) thay đổi xử lý âm thanh có thể khiến giảm mức tái tạo loa thứ nhất đối với loa thứ nhất của thiết bị âm thanh thứ nhất và có thể khiến giảm mức tái tạo loa thứ hai đối với loa thứ hai của thiết bị âm thanh thứ hai. Theo một số ví dụ như vậy, việc giảm mức tái tạo loa có thể tương đối lớn hơn đối với thiết bị âm thanh có khoảng cách lớn hơn so với vị trí hiện tại được ước lượng của người (hoặc với vị trí micrô gần nhất với vị trí hiện tại được ước lượng của người đó).

Tuy nhiên, các tác giả sáng chế dự tính nhiều kiểu thay đổi xử lý âm thanh có thể được thực hiện trong một số trường hợp. Theo một số phương án thực hiện, một hoặc nhiều kiểu thay đổi xử lý âm thanh có thể bao gồm bước thay đổi quy trình kết xuất để làm lệch kết xuất tín hiệu âm thanh khỏi vị trí hiện tại được ước lượng của người đó (hoặc từ vị trí micrô gần nhất với vị trí hiện tại được ước lượng của người đó).

Theo một số phương án thực hiện, một hoặc nhiều kiểu thay đổi xử lý âm thanh có thể bao gồm bước sửa đổi phổ. Ví dụ, bước sửa đổi phổ có thể bao gồm bước giảm mức dữ liệu âm thanh trong băng tần số từ 500 Hz đến 3 Khz. Trong các ví dụ khác, bước sửa đổi phổ có thể bao gồm bước giảm mức dữ liệu âm thanh trong băng tần số có tần số tối đa cao hơn và/hoặc tần số tối thiểu thấp hơn. Theo một số phương án thực hiện, một hoặc nhiều kiểu thay đổi xử lý âm thanh có thể bao gồm bước chèn ít nhất một khe vào ít nhất một băng tần số đã chọn của tín hiệu phát lại âm thanh.

Theo một số phương án thực hiện, bước xác định một hoặc nhiều kiểu thay đổi xử lý âm thanh có thể dựa trên việc tối ưu hóa hàm chi phí mà dựa, ít nhất một phần, vào ước lượng tỷ lệ tín hiệu trên tiếng vọng. Trong một số trường hợp, hàm chi phí có thể dựa, ít nhất một phần, vào hiệu suất kết xuất.

Theo ví dụ này, khối 275 bao gồm bước khiến cho một hoặc nhiều kiểu thay đổi xử lý âm thanh sẽ được áp dụng. Trong một số trường hợp, khối 275 có thể bao gồm bước áp dụng một hoặc nhiều kiểu thay đổi xử lý âm thanh, bằng một hoặc nhiều thiết bị điều khiển quy trình xử lý âm thanh trong môi trường âm thanh. Trong các trường hợp khác, khối 275 có thể bao gồm bước khiến cho (ví dụ, qua lệnh hoặc tín hiệu điều khiển từ bộ quản lý phiên âm thanh) một hoặc nhiều kiểu thay đổi xử lý âm thanh được một hoặc nhiều thiết bị khác của môi trường âm thanh áp dụng.

Một số phương án thực hiện của phương pháp 250 có thể bao gồm bước lựa chọn ít nhất một micrô theo một hoặc nhiều khía cạnh của thông tin ngữ cảnh. Theo một số phương án thực hiện như vậy, phương pháp 250 có thể bao gồm bước lựa chọn ít nhất một micrô theo độ gần hiện tại được ước lượng của người đó với một hoặc nhiều vị trí micrô. Một số phương án thực hiện của phương pháp 250 có thể bao gồm bước lựa chọn ít nhất một micrô theo ước lượng về vùng người dùng. Theo một số phương án thực hiện như vậy, phương pháp 250 có thể bao gồm bước áp dụng, ít nhất một phần, chức năng trợ lý ảo theo tín hiệu micrô nhận được từ (các) micrô đã chọn. Theo một số phương án thực hiện như vậy, phương pháp 250 có thể bao gồm bước cung cấp chức năng hội nghị từ xa dựa, ít nhất một phần, vào tín hiệu micrô nhận được từ (các) micrô đã chọn.

Một số phương án cung cấp hệ thống (bao gồm hai hoặc nhiều thiết bị, ví dụ, thiết bị âm thanh thông minh) được tạo cấu hình để áp dụng kỹ thuật kết xuất và ánh xạ, và được tạo cấu hình để sử dụng phần mềm hoặc các dạng logic khác (ví dụ, để bao gồm phần tử hệ thống áp dụng logic) để thay đổi kỹ thuật xử lý âm thanh (ví dụ, giảm âm lượng một, hai hoặc nhiều loa gần nhất). Logic có thể áp dụng bộ giám sát, như thiết bị được tạo cấu hình để áp dụng bộ quản lý phiên âm thanh, theo một số ví dụ, có thể chạy riêng với các phần tử hệ thống được tạo cấu hình để kết xuất.

Fig.3A là sơ đồ khối của hệ thống được tạo cấu hình để thực hiện điều khiển kết xuất và nghe hoặc thu nhận logic trên nhiều thiết bị. Số lượng, kiểu và cách sắp xếp các phần tử được thể hiện trên các hình vẽ từ Fig.3A đến Fig.3C, cũng như các hình vẽ khác được mô tả chỉ là ví dụ. Các phương án thực hiện khác có thể bao gồm nhiều hơn, ít hơn và/hoặc các loại phần tử khác nhau. Ví dụ, các phương án thực hiện khác có thể bao gồm nhiều hơn ba thiết bị âm thanh, các loại thiết bị âm thanh khác nhau, v.v.

Các môđun được thể hiện trên các hình vẽ từ Fig.3A đến Fig.3C, cũng như các môđun khác được thể hiện và mô tả trong sáng chế này, có thể được áp dụng qua phần cứng, phần mềm, firmware, v.v., tùy thuộc vào ví dụ cụ thể. Theo một số phương án thực hiện, một hoặc nhiều môđun (trong một số trường hợp có thể được gọi ở đây là “phần tử”) có thể được thực hiện qua hệ thống điều khiển như hệ thống điều khiển 210 được mô tả ở trên dựa vào hình vẽ trên Fig.2A. Theo một số ví dụ như vậy, một hoặc nhiều môđun được mô tả có thể được áp dụng theo phần mềm mà được thực thi bằng một hoặc nhiều hệ thống điều khiển như vậy.

Các phần tử trên Fig.3A bao gồm:

Các thiết bị âm thanh 302, 303 và 304 (theo một số ví dụ có thể là thiết bị âm thanh thông minh). Theo ví dụ này, mỗi thiết bị âm thanh 302, 303 và 304 bao gồm ít nhất một loa và ít nhất một micrô;

- Phần tử 300 là một dạng nội dung, bao gồm dữ liệu âm thanh, sẽ được phát trên một hoặc nhiều thiết bị âm thanh 302, 303 và 304. Nội dung 300 có thể là nội dung tuyến tính hoặc tương tác, tùy thuộc vào phương án thực hiện cụ thể;
- Môđun 301 được tạo cấu hình để xử lý âm thanh, bao gồm nhưng không giới hạn ở việc kết xuất theo logic kết xuất. Ví dụ, theo một số phương án, môđun 301 có thể được tạo cấu hình đơn giản để sao chép âm thanh của nội dung 300 (ví dụ, đơn âm hoặc âm thanh nổi) cho cả ba thiết bị âm thanh 302, 303 và 304 như nhau. Theo một số phương án thực hiện khác, một hoặc nhiều thiết bị âm thanh 302, 303 và 304 có thể được tạo cấu hình để áp dụng chức năng xử lý âm thanh, bao gồm nhưng không giới hạn ở chức năng kết xuất;
- Phần tử 305 đại diện cho các tín hiệu được phân phối cho các thiết bị âm thanh 302, 303 và 304. Theo một số ví dụ, tín hiệu 305 có thể là, hoặc có thể bao gồm, tín hiệu cấp cho loa. Như đã lưu ý ở trên, theo một số phương án thực hiện, chức năng của môđun 301 có thể được áp dụng qua một hoặc nhiều thiết bị âm thanh 302, 303 và 304, trong trường hợp đó, tín hiệu 305 có thể là tín hiệu cục bộ cho một hoặc nhiều thiết bị âm thanh 302, 303 và 304. Tuy nhiên, các tín hiệu này được thể hiện trên Fig.3A dưới dạng tập hợp các tín hiệu cấp cho loa, vì một số phương án (ví dụ, như được mô tả dưới đây dựa vào Fig.4) thực hiện kỹ thuật chặn cuối cùng đơn giản hoặc xử lý sau các tín hiệu 305;

- Phần tử 306 đại diện cho tín hiệu micrô thô được thu nhận bằng micrô của các thiết bị âm thanh 302, 303 và 304.
- Môđun 307 được tạo cấu hình để sử dụng logic xử lý tín hiệu micrô và, theo một số ví dụ, logic thu tín hiệu micrô. Vì trong ví dụ này, mỗi trong số các thiết bị âm thanh 302, 303 và 304 có một hoặc nhiều micrô, nên các tín hiệu thô thu được 306 được xử lý bằng môđun 307. Theo một số phương án thực hiện, như ở đây, môđun 307 có thể được tạo cấu hình để áp dụng chức năng khử tiếng vọng và/hoặc chức năng phát hiện tiếng vọng.
- Phần tử 308 biểu diễn các tín hiệu tham chiếu tiếng vọng cục bộ và/hoặc toàn cục mà được cung cấp bằng môđun 301 cho môđun 307. Theo ví dụ này, môđun 307 được tạo cấu hình để áp dụng chức năng khử tiếng vọng và/hoặc phát hiện tiếng vọng theo tín hiệu tham chiếu tiếng vọng cục bộ và/hoặc toàn cục 308. Theo một số phương án thực hiện, việc xử lý thu micrô và/hoặc xử lý tín hiệu micrô thô có thể được phân phối bằng mạch logic phát hiện và khử tiếng vọng cục bộ trên mỗi thiết bị trong số các thiết bị âm thanh 302, 303 và 304. Phương án thực hiện cụ thể về việc thu và xử lý thu dữ liệu là không quan trọng đối với khái niệm tính toán và hiệu tác động của mọi thay đổi đối với việc kết xuất trên SER tổng thể và tính hiệu quả của quy trình xử lý thu và logic;
- Môđun 309 là phần tử hệ thống thực hiện trộn hoặc kết hợp tổng thể của âm thanh thu được (ví dụ, để khiến cho âm thanh mong muốn được cảm nhận là phát ra từ một vị trí cụ thể hoặc vị trí rộng). Theo một số phương án thực hiện, môđun 307 cũng có thể cung cấp chức năng trộn của phần tử 309; và
- Môđun 310 là phần tử hệ thống thực hiện khía cạnh xử lý âm thanh phát hiện được cuối cùng để đưa ra quyết định nào đó về những gì đang được nói hoặc nếu có hoạt động quan tâm diễn ra trong môi trường âm thanh. Ví dụ, môđun 310 có thể cung cấp chức năng nhận dạng lời nói tự động (automatic speech recognition - ASR), chức năng cảm biến kiểu và/hoặc mức tiếng ồn xung quanh, ví dụ, đối với ngữ cảnh liên quan đến con người đang làm việc trong môi trường âm thanh, mức tiếng ồn tổng thể trong môi trường âm thanh, v.v. Theo một số phương án thực hiện, một số hoặc tất cả chức năng của môđun 310 có thể được áp dụng bên ngoài môi trường âm thanh trong đó các thiết bị âm thanh 302, 303 và 304 nằm,

ví dụ, trong một hoặc nhiều thiết bị (ví dụ, một hoặc nhiều máy chủ) của nhà cung cấp dịch vụ dựa trên đám mây.

Fig.3B là sơ đồ khối của hệ thống theo phương án thực hiện khác được mô tả. Trong ví dụ này, hệ thống được thể hiện trên Fig.3B bao gồm các phần tử của hệ thống trên Fig.3A và mở rộng hệ thống trên Fig.3A để bao gồm chức năng theo một số phương án được mô tả. Hệ thống trên Fig.3B bao gồm các phần tử áp dụng các khía cạnh NGŨ CẢNH, QUYẾT ĐỊNH và HÀNH ĐỘNG KẾT XUẤT, như được áp dụng cho hệ thống âm thanh được phân tán đang hoạt động. Theo một số ví dụ, việc phản hồi đối với các phần tử để thực hiện các khía cạnh của NGŨ CẢNH, QUYẾT ĐỊNH và HÀNH ĐỘNG KẾT XUẤT có thể làm tăng độ tin cậy nếu có hoạt động (ví dụ, phát hiện lời nói), hoặc khả năng để làm giảm theo cách tin cậy sự nhận biết hoạt động (khả năng hoạt động thấp) và do đó đưa quy trình xử lý âm thanh về trạng thái ban đầu.

Các phần tử trên Fig.3B bao gồm:

- Môđun 351 là phần tử hệ thống đại diện cho (và áp dụng) bước NGŨ CẢNH, ví dụ, để thu được chỉ báo về vị trí có thể được mong muốn để phát hiện âm thanh tốt hơn (ví dụ, tăng tỷ lệ lời nói trên tiếng vọng tại một hoặc nhiều micrô), và khả năng hoặc cảm giác mà con người muốn nghe (ví dụ, xác suất lời nói đó, như từ đánh thức hoặc lệnh, có thể được một hoặc nhiều micrô ghi lại). Trong ví dụ này, các môđun 351 và 353 được thực hiện qua hệ thống điều khiển, là hệ thống điều khiển 210 trên Fig.2A trong trường hợp này. Theo một số phương án thực hiện, các khối 301 và 307 cũng có thể được thực hiện bằng hệ thống điều khiển, trong một số trường hợp, hệ thống này có thể là hệ thống điều khiển 210. Theo một số phương án thực hiện, các khối 356, 357 và 358 cũng có thể được thực hiện bằng hệ thống điều khiển, trong một số trường hợp có thể là hệ thống điều khiển 210;
- Phần tử 352 đại diện cho đường dẫn phản hồi với môđun 351. Phản hồi 352 được cung cấp bằng môđun 310 trong ví dụ này. Theo một số phương án, phản hồi 352 có thể tương ứng với kết quả xử lý âm thanh (như xử lý âm thanh cho ASR) từ việc thu các tín hiệu micrô có thể liên quan đến bước xác định ngữ cảnh - ví dụ, khả năng phát hiện sớm hoặc yếu từ đánh thức hoặc phát hiện thấp hoạt động lời nói có thể được dùng để bắt đầu nâng cao sự tin cậy hoặc nhận biết về ngữ cảnh

muốn cải thiện khả năng nghe (ví dụ, để tăng tỷ lệ lời nói trên tiếng vọng tại một hoặc nhiều micrô);

- Modul 353 là phần tử hệ thống trong đó (hoặc nhờ đó) quyết định được tạo ra liên quan đến thiết bị âm thanh nào sẽ thay đổi quy trình xử lý âm thanh và số lượng thiết bị đó. Tùy thuộc vào phương án thực hiện cụ thể, modul 353 có thể sử dụng hoặc không sử dụng thông tin thiết bị âm thanh cụ thể, như loại và/hoặc khả năng của thiết bị âm thanh (ví dụ, khả năng của loa, khả năng khử tiếng vọng, v.v.), hướng của thiết bị âm thanh, v.v. Như được mô tả theo một số ví dụ dưới đây, (các) quy trình ra quyết định của modul 353 có thể rất khác đối với thiết bị tai nghe so với loa thông minh hoặc loa khác;
- Phần tử 354 là đầu ra của modul 353, theo ví dụ này, là tập hợp các chức năng điều khiển, được biểu diễn trên Fig.3B là các giá trị f_n , đối với các khối kết xuất riêng qua đường dẫn điều khiển 355, đường dẫn này cũng có thể được gọi là đường dẫn tín hiệu 355. Tập hợp chức năng điều khiển có thể được phân phối (ví dụ, qua cuộc truyền không dây) sao cho đường dẫn tín hiệu 355 này là cục bộ đối với môi trường âm thanh. Theo ví dụ này, các chức năng điều khiển được cung cấp cho các modul 356, 357 và 358; và
- Các modul 356, 357 và 358 là các phần tử hệ thống được tạo cấu hình để thực hiện thay đổi trong xử lý âm thanh, có khả năng bao gồm nhưng không giới hạn ở kết xuất đầu ra (khía cạnh KẾT XUẤT theo một số phương án). Theo ví dụ này, các modul 356, 357 và 358 được điều khiển kích hoạt bằng các chức năng điều khiển (trong ví dụ này, các giá trị f_n) của đầu ra 354. Theo một số phương án thực hiện, chức năng của các modul 356, 357 và 358 có thể được thực hiện qua khối 301.

Trong phương án trên Fig.3B và các phương án thực hiện khác, chu kỳ phản hồi chính xác có thể xảy ra. Nếu đầu ra 352 của phần tử 310 (mà có thể áp dụng kỹ thuật nhận dạng lời nói tự động hay ASR trong một số trường hợp) phát hiện lời nói, thậm chí yếu (ví dụ, có độ tin cậy thấp), theo một số ví dụ, thì phần tử NGŨ CẢNH 351 có thể ước lượng vị trí dựa trên (các) micrô nào thu được âm thanh trong môi trường âm thanh (ví dụ, (các) micrô nào có hầu hết năng lượng khác tiếng vọng). Theo một số ví dụ như vậy, khối QUYẾT ĐỊNH 353 có thể lựa chọn một, hai, ba hoặc nhiều loa của môi trường

âm thanh và có thể kích hoạt giá trị nhỏ liên quan đến sự thay đổi khi kết xuất (ví dụ, $f_n = 0,25$). Trong trường hợp giảm âm thanh 20 dB tổng thể, giá trị này sau đó sẽ thực hiện giảm âm lượng ở (các) thiết bị được chọn khoảng 5 dB để có thể đủ nhận biết đối với người nghe bình thường. (Các) mức giảm có thể ít nhận biết được hơn khi được kết hợp với hằng số thời gian và/hoặc việc phát hiện sự kiện, và với các loa khác của môi trường âm thanh tái tạo nội dung tương tự. Theo một ví dụ, thiết bị này có thể là thiết bị âm thanh 303 (thiết bị âm thanh gần người 311 đang nói nhất) được giảm âm lượng. Trong các ví dụ khác, cả thiết bị âm thanh 302 và 303 đều có thể được giảm âm lượng, trong một số trường hợp, có lượng khác nhau (ví dụ, phụ thuộc vào độ gần được ước lượng với người 311). Trong các ví dụ khác, tất cả các thiết bị âm thanh 302, 303 và 304 có thể được giảm âm lượng, trong một số trường hợp, có lượng khác nhau. Do giảm mức phát lại của một hoặc nhiều loa của một hoặc nhiều thiết bị âm thanh 302, 303 và 304, tỷ lệ lời nói trên tiếng vọng có thể được tăng lên ở một hoặc nhiều micrô gần người 311 (ví dụ, một hoặc nhiều micrô của thiết bị âm thanh 303). Do đó, hệ thống giờ đây có thể “nghe” người 311 tốt hơn nếu người 311 tiếp tục nói (ví dụ, lặp lại từ đánh thức hoặc để nói lệnh). Theo một số phương án thực hiện như vậy, trong khoảng thời gian tiếp theo (ví dụ, trong vài giây tiếp theo) và, trong một số trường hợp liên tục, hệ thống (ví dụ, bộ quản lý phiên âm thanh được áp dụng, ít nhất một phần, qua các khối 351 và 353) có thể nhanh chóng có xu hướng tắt âm lượng của một hoặc nhiều loa gần người 311, ví dụ, bằng cách lựa chọn $f_2 = 1$.

Fig.3C là sơ đồ khối của một phương án được tạo cấu hình để áp dụng mạng cân bằng năng lượng theo một ví dụ. Fig.3C là sơ đồ khối của hệ thống bao gồm các phần tử của hệ thống trên Fig.3B và mở rộng hệ thống trên Fig.3B để bao gồm các phần tử (ví dụ, phần tử 371) để thực hiện việc bù năng lượng (ví dụ, để ‘tăng âm lượng các thiết bị khác một chút’).

Theo một số ví dụ, thiết bị được tạo cấu hình để quản lý phiên âm thanh (bộ quản lý phiên âm thanh) của hệ thống trên Fig.3C (hoặc hệ thống như hệ thống trên Fig.3C) có thể ước lượng năng lượng tạo dải ở người nghe (311) mà bị mất do ảnh hưởng của quy trình xử lý âm thanh (ví dụ, giảm mức của một hoặc nhiều loa đã chọn (ví dụ, loa của thiết bị âm thanh nhận tín hiệu điều khiển trong đó $f_n > 0$) mà được áp dụng để tăng tỷ lệ lời nói trên tiếng vọng ở một hoặc nhiều micrô). Sau đó, bộ quản lý phiên âm thanh

có thể áp dụng việc tăng cường mức và/hoặc dạng cân bằng năng lượng khác nào đó cho các loa khác của môi trường âm thanh để bù các thay đổi xử lý âm thanh dựa trên SER.

Thông thường, khi kết xuất nội dung có liên quan và có các thành phần âm thanh được làm tương quan hoặc có phổ tương tự được tái tạo bằng nhiều loa của môi trường âm thanh (ví dụ đơn giản là đơn âm) thì có thể không cần cân bằng năng lượng nhiều. Ví dụ, nếu có 3 loa trong môi trường âm thanh với khoảng cách theo tỷ lệ từ 1 đến 2, 1 là loa gần nhất, thì việc giảm âm lượng loa gần nhất 6 dB nếu có nội dung giống hệt nhau được tái tạo bằng các loa sẽ chỉ có ảnh hưởng từ 2 đến 3 dB. Và việc tắt loa gần nhất có thể chỉ có ảnh hưởng tổng thể từ 3 đến 4 dB về âm thanh tại vị trí người nghe.

Trong các tình huống phức tạp hơn (ví dụ, chèn các khe hoặc điều khiển theo không gian), theo một số ví dụ, dạng bảo toàn năng lượng và sự liên tục về cảm nhận có thể là sự cân bằng năng lượng nhiều yếu tố hơn.

Trên Fig.3C, (các) phần tử để áp dụng NGŨ CẢNH có thể, theo một số ví dụ, chỉ là mức âm thanh (nghịch đảo của độ gần) của việc phát hiện yếu tố đánh thức. Nói cách khác, một ví dụ về việc xác định NGŨ CẢNH có thể dựa trên mức của bất kỳ tiếng nói từ đánh thức nào được phát hiện qua tiếng vọng phát hiện được. Các phương pháp như vậy có thể có hoặc có thể không thực sự bao gồm bước xác định tỷ lệ lời nói trên tiếng vọng, tùy thuộc vào phương án thực hiện cụ thể. Tuy nhiên, việc phát hiện và đánh giá đơn giản mức của tiếng nói từ đánh thức được phát hiện tại mỗi trong số nhiều vị trí micro có thể cung cấp mức NGŨ CẢNH đủ trong một số ví dụ.

Ví dụ về các phương pháp được thực hiện bằng các phần tử hệ thống để áp dụng NGŨ CẢNH (ví dụ, trong hệ thống trên Fig.3C) có thể bao gồm, nhưng không giới hạn ở, những yếu tố sau:

- Khi phát hiện một phần của từ đánh thức, độ gần với các thiết bị âm thanh được trang bị micro có thể được suy ra từ độ tin cậy của từ đánh thức. Việc định thời của tiếng nói từ đánh thức cũng có thể được suy ra từ độ tin cậy của từ đánh thức; và
- Hoạt động âm thanh nào đó có thể được phát hiện ngoài việc khử và triệt tiếng vọng được áp dụng cho tín hiệu micro thô. Một số phương án thực hiện có thể sử dụng tập hợp các mức năng lượng và phân loại để xác định xem xác suất hoạt

động của âm thanh là hoạt động giọng nói (Phát hiện hoạt động giọng nói). Quy trình này có thể xác định độ tin cậy hoặc xác suất hoạt động giọng nói. Vị trí của giọng nói có thể dựa trên xác suất của micrô tốt nhất cho các tình huống tương tác tương tự. Ví dụ, thiết bị áp dụng bộ quản lý phiên âm thanh có thể biết trước rằng một thiết bị âm thanh được trang bị micrô ở gần người dùng nhất, như thiết bị để bàn ở hoặc ở gần vị trí chung với người dùng, thay vì thiết bị được lắp trên tường không gần vị trí chung với người dùng.

Một phương án làm ví dụ về phần tử hệ thống để áp dụng QUYẾT ĐỊNH (ví dụ, trong hệ thống trên Fig.3C) là phần tử được tạo cấu hình để xác định giá trị độ tin cậy liên quan đến hoạt động giọng nói và để xác định thiết bị âm thanh được trang bị micrô nào là gần nhất.

Trong hệ thống trên Fig.3C (và các phương án khác), lượng (các) thay đổi xử lý âm thanh cần áp dụng để tăng SER tại một vị trí có thể là hàm của khoảng cách và độ tin cậy liên quan đến hoạt động giọng nói.

Ví dụ về các phương pháp thực hiện KẾT XUẤT (ví dụ, trong hệ thống trên Fig.3C) bao gồm:

Giảm âm lượng một dB; và/hoặc

Cân bằng (EQ) băng tần lời nói (ví dụ, như được mô tả bên dưới dựa vào Fig.4); và/hoặc

Điều chế thời gian của quy trình thay đổi kết xuất (như được mô tả dựa vào hình vẽ trên Fig.5); và/hoặc

Sử dụng kỹ thuật cắt lát thời gian theo thời gian hoặc điều chỉnh để tạo ra (ví dụ, chèn vào nội dung âm thanh) "khe" hoặc khoảng thời gian đầu ra tần số thời gian thừa thớt hơn đủ để thu được sự thoáng qua của âm thanh quan tâm. Một số ví dụ được ô tả dưới đây liên quan đến Fig.9

Fig.4 là biểu đồ minh họa các ví dụ về xử lý âm thanh có thể làm tăng tỷ lệ lời nói trên tiếng vọng tại một hoặc nhiều micrô của môi trường âm thanh. Đồ thị trên Fig.4 cung cấp các ví dụ về sự sửa đổi phổ. Trên Fig.4, bước sửa đổi phổ bao gồm bước giảm mức tần số đã biết tương ứng với lời nói, trong các ví dụ này là các tần số trong dải từ 200 Hz đến 10 KHz (ví dụ, trong khoảng 5% hoặc 10% của tần số cao hơn và/hoặc tần số thấp hơn của dải). Các ví dụ khác có thể bao gồm bước giảm mức tần số trong băng

tần số khác, ví dụ, giữa khoảng 500 Hz và 3 KHz (ví dụ, trong khoảng 5% hoặc 10% của tần số cao hơn và/hoặc thấp hơn của dải). Theo một số phương án thực hiện, các tần số nằm ngoài dải này có thể được tái tạo ở mức cao hơn, để ít nhất bù một phần cho việc giảm âm lượng do sửa đổi phổ gây ra.

Các phần tử trên Fig.4 bao gồm:

601: đường cong biểu diễn EQ phẳng;

602: đường cong biểu diễn sự suy giảm một phần của dải tần số được biểu thị. Sự suy giảm một phần như vậy có thể có khả năng nhận thấy tương đối thấp nhưng vẫn có thể có ảnh hưởng có lợi đến việc phát hiện lời nói; và

603: đường cong biểu diễn sự suy giảm lớn hơn đáng kể của dải tần số được biểu thị. Sự sửa đổi phổ như được biểu thị bằng đường cong 603 có thể có ảnh hưởng lớn đến việc nghe lời nói. Trong một số trường hợp, việc sửa đổi phổ tích cực như được biểu diễn bằng đường cong 603 có thể cung cấp giải pháp khác để giảm đáng kể mức của tất cả các tần số.

Theo một số ví dụ, bộ quản lý phiên âm thanh có thể khiến cho các thay đổi xử lý âm thanh tương ứng với việc sửa đổi phổ biến đổi theo thời gian, như chuỗi được biểu thị bằng các đường cong 601, 602 và 603.

Theo một số ví dụ, một hoặc nhiều sửa đổi phổ có thể được dùng trong ngữ cảnh có các thay đổi xử lý âm thanh khác, như ngữ cảnh thay đổi kết xuất để “làm lệch” âm thanh tái tạo cách xa vị trí, như văn phòng, phòng ngủ, em bé đang ngủ, v.v. (Các) sửa đổi phổ được dùng liên quan đến hiện tượng lệch như vậy, chẳng hạn, có thể làm giảm mức trong dải tần số âm trầm, ví dụ, trong dải từ 20 Hz đến 250 Hz.

Fig.5 là biểu đồ minh họa kiểu xử lý âm thanh khác mà có thể làm tăng tỷ lệ lời nói trên tiếng vọng tại một hoặc nhiều micrô của môi trường âm thanh. Trong ví dụ này, trục tung biểu thị các giá trị "f" từ 0 đến 1 và trục hoành biểu thị thời gian, tính bằng giây. Fig.5 là sơ đồ quỹ đạo (được biểu thị bằng đường cong 701) theo thời gian của quy trình kích hoạt hiệu ứng kết xuất. Theo một số ví dụ, một hoặc nhiều môđun 356, 357 hoặc 358 có thể sử dụng kiểu xử lý âm thanh được mô tả trên Fig.5. Theo ví dụ này, sự không đối xứng của hằng số thời gian (được biểu thị bằng đường cong 701) biểu thị rằng hệ thống điều chế giá trị điều khiển (f_n) trong thời gian ngắn (ví dụ, từ 100 mili giây đến 1 giây) nhưng giảm chậm hơn nhiều (ví dụ, 10 giây trở lên) từ giá trị f_n (cũng được

xác định là giá trị 703) trở về không. Theo một số ví dụ, khoảng thời gian từ 2 giây đến N giây có thể là nhiều giây, ví dụ, trong khoảng từ 4 đến 10 giây.

Như được thể hiện trên Fig.5 là đường cong kích hoạt thứ hai 702 có dạng bậc thang với giá trị lớn nhất bằng f_n trong ví dụ này. Theo phương án thực hiện này, các bước tăng tương ứng với các thay đổi mức đột ngột của chính nội dung, ví dụ, tốc độ bắt đầu nói hoặc tốc độ âm tiết.

Như đã lưu ý ở trên, theo một số phương án thực hiện, việc cắt lát thời gian theo thời gian hoặc điều chỉnh tần số có thể tạo ra (ví dụ, bằng cách chèn các khe vào nội dung âm thanh) "khe" hoặc khoảng thời gian đầu ra tần số thời gian thừa thớt đủ để thu được sự cảm nhận thoáng qua về âm thanh quan tâm (ví dụ, tăng hoặc giảm mức độ "khe" của nội dung âm thanh và cảm nhận của chúng).

Fig.6 minh họa kiểu xử lý âm thanh khác có thể làm tăng tỷ lệ lời nói trên tiếng vọng tại một hoặc nhiều micro của môi trường âm thanh. Fig.6 là một ví dụ về ảnh phổ của tín hiệu phát lại âm thanh sửa đổi, trong đó các khe cưỡng bức được chèn vào theo một ví dụ. Cụ thể hơn, để tạo ra ảnh phổ trên Fig.6, các khe cưỡng bức G1, G2 và G3 được chèn vào các băng tần số của tín hiệu phát lại, do đó tạo ra tín hiệu phát lại âm thanh sửa đổi. Trong ảnh phổ thể hiện trên Fig.6, vị trí dọc theo trục hoành biểu thị thời gian và vị trí dọc theo trục tung biểu thị tần số của nội dung của tín hiệu phát lại âm thanh sửa đổi tại một thời điểm.

Mật độ dấu chấm trong mỗi vùng nhỏ (mỗi vùng như vậy tập trung tại một điểm có tọa độ dọc và ngang) biểu thị năng lượng của nội dung tín hiệu phát lại âm thanh sửa đổi ở tần số và thời điểm tương ứng (các vùng dày đặc hơn biểu thị nội dung có năng lượng lớn hơn, và các vùng ít dày đặc hơn biểu thị nội dung có năng lượng thấp hơn). Do đó, khe G1 xuất hiện tại một thời điểm (ví dụ, trong một khoảng thời gian) sớm hơn thời điểm mà (hoặc khoảng thời gian trong đó) khe G2 hoặc G3 xuất hiện tại đó, và khe G1 đã được chèn vào băng tần số cao hơn băng tần số trong đó khe G2 hoặc G3 đã được chèn.

Việc đưa khe cưỡng bức vào tín hiệu phát lại khác với hoạt động của thiết bị đơn công trong đó thiết bị tạm dừng luồng nội dung phát lại (ví dụ, để nghe người dùng và môi trường của người dùng rõ hơn). Việc đưa các khe cưỡng bức vào tín hiệu phát lại theo một số phương án được mô tả có thể được tối ưu hóa để giảm đáng kể (hoặc loại

bỏ) khả năng nhận biết nhiễu tạo ra từ các khe đưa vào trong quá trình phát lại, tốt hơn là các khe cường bức không có hoặc có ảnh hưởng nhận biết được tối thiểu đến người dùng, nhưng sao cho tín hiệu đầu ra của micrô trong môi trường phát lại là biểu thị của các khe cường bức (ví dụ, vì vậy các khe có thể được sử dụng để thực hiện phương pháp nghe rộng khắp). Bằng cách sử dụng các khe cường bức đã được đưa vào theo một số phương án được mô tả, hệ thống nghe rộng khắp có thể giám sát âm thanh không phát lại (ví dụ, âm thanh biểu thị hoạt động và/hoặc tiếng ồn nền trong môi trường phát lại) ngay cả khi không sử dụng bộ khử tiếng vọng âm thanh.

Theo một số ví dụ, các khe có thể được chèn vào đầu ra phổ thời gian từ một kênh, việc này có thể tạo ra cảm giác thừa thớt về khả năng nghe cải thiện để “nghe qua các khe”.

Fig.7 là biểu đồ minh họa kiểu xử lý âm thanh khác có thể làm tăng tỷ lệ lời nói trên tiếng vọng tại một hoặc nhiều micrô của môi trường âm thanh. Theo phương án này, các thay đổi trong xử lý âm thanh bao gồm nén dải động.

Ví dụ này bao gồm bước chuyển đổi giữa hai cực giới hạn dải động. Trong một trường hợp, được biểu diễn bằng đường cong 801, bộ quản lý phiên âm thanh không áp dụng kỹ thuật điều khiển dải động, trong khi trong trường hợp còn lại, được biểu diễn bằng đường cong 802, bộ quản lý phiên âm thanh khiến cho bộ giới hạn tương đối tích cực sẽ được áp dụng. Bộ giới hạn tương ứng với đường cong 802 có thể giảm các đỉnh của đầu ra âm thanh từ 10 dB trở lên. Theo một số ví dụ, tỷ lệ nén có thể không quá 3:1. Theo một số phương án thực hiện, đường cong 802 (hoặc đường cong nén dải động khác) có thể bao gồm chỗ uốn ở -20 dB, hoặc khoảng -20 dB (ví dụ, trong ± 1 dB, trong ± 2 dB, trong ± 3 dB, v.v.) từ đầu ra đỉnh của thiết bị.

Tiếp theo, sáng chế mô tả ví dụ khác về một phương án của các phần tử hệ thống để áp dụng kỹ thuật KẾT XUẤT (các thay đổi xử lý âm thanh có tác dụng tăng tỷ lệ lời nói trên tiếng vọng ở một hoặc nhiều micrô, ví dụ, trong hệ thống trên Fig.3B hoặc trên Fig.3C). Theo phương án này, kỹ thuật cân bằng năng lượng được thực hiện. Như đã lưu ý ở trên, theo một ví dụ đơn giản, bộ quản lý phiên âm thanh có thể đánh giá năng lượng tạo dải của âm thanh tại vị trí hoặc vùng của người nghe bị mất do ảnh hưởng của sự thay đổi xử lý âm thanh khác để tăng SER tại một hoặc nhiều micrô của môi trường

âm thanh. Sau đó, bộ quản lý phiên âm thanh có thể thêm phần tăng áp vào các loa khác để bù lại phần năng lượng đã mất này tại vị trí hoặc vùng của người nghe.

Thông thường, khi sáng chế kết xuất nội dung có liên quan và có các thành phần tương quan hoặc tương tự về mặt phổ trong nhiều thiết bị (ví dụ đơn giản là đơn âm) thì có thể không cần phải làm gì nhiều. Ví dụ, nếu có 3 loa có các khoảng cách theo tỷ lệ từ 1 đến 2, 1 là loa gần nhất, thì việc giảm âm lượng loa gần nhất 6 dB (nếu nội dung giống nhau được tái tạo bằng các loa) sẽ chỉ có tác động từ 2 dB đến 3 dB. Và việc tắt loa gần nhất có thể chỉ ảnh hưởng chung từ 3 dB đến 4 dB đối với âm thanh tại vị trí của người nghe.

Tiếp theo, sáng chế mô tả các khía cạnh của các phương án bổ sung.

1. CÁC HỆ SỐ BẬC HAI TRONG VIỆC ĐỊNH NGHĨA ‘GẦN NHẤT’

Như hai ví dụ sau đây sẽ minh họa, số đo “độ gần” hoặc “gần nhất” có thể không phải là số đo khoảng cách đơn giản, mà thay vào đó có thể là một loại vô hướng liên quan đến tỷ lệ lời nói trên tiếng vọng được ước lượng. Nếu các thiết bị âm thanh của môi trường âm thanh không giống nhau, thì mỗi thiết bị âm thanh được trang bị loa có thể có cách ghép khác nhau từ (các) loa của nó với (các) micrô của chính nó, có ảnh hưởng mạnh đến mức vọng âm trong tỷ lệ. Ngoài ra, các thiết bị âm thanh có thể có các cách bố trí micrô khác nhau mà có sự phù hợp tương đối nhiều hoặc tương đối ít để nghe, ví dụ, để phát hiện âm thanh từ một hướng cụ thể, để phát hiện âm thanh trong hoặc từ một vị trí cụ thể của môi trường âm thanh. Do đó, theo một số phương án thực hiện, phép tính (QUYẾT ĐỊNH) có thể ảnh hưởng nhiều hơn đến độ gần và tính thuận nghịch của thính giác.

Fig.8 là sơ đồ của một ví dụ trong đó thiết bị âm thanh được giảm âm lượng có thể không phải là thiết bị âm thanh gần nhất với người đang nói. Trong ví dụ này, thiết bị âm thanh 802 tương đối gần với người nói 100 hơn so với thiết bị âm thanh 805. Theo một số ví dụ, trong các tình huống như thể hiện trên Fig.8, bộ quản lý phiên âm thanh có thể xem xét các đặc tính SER cơ bản và thiết bị âm thanh khác nhau và giảm âm lượng (các) thiết bị có tỷ lệ chi phí/lợi ích tốt nhất về tác động của việc giảm đầu ra đối với dạng biểu diễn âm thanh, so với lợi ích giảm âm lượng đầu ra để ghi lại tốt hơn lời nói của người 101.

Fig.8 thể hiện một ví dụ trong đó có thể có sự phức tạp và hữu ích trong phép đo nhiều chức năng của “gần nhất”. Trong trường hợp này, có người 101 phát ra âm thanh (lời nói 102) mà bộ quản lý phiên âm thanh được tạo cấu hình để thu và hai thiết bị âm thanh 802 và 805, cả hai đều có loa (806 và 804) và micrô (803 và 807). Xác định rằng micrô 803 gần loa 804 trên thiết bị âm thanh 802 để gần với người 101 hơn, nên có thể không giảm âm lượng loa của thiết bị này mà có thể đạt được SER khả thi. Trong ví dụ này, micrô 807 trên thiết bị âm thanh 805 được tạo cấu hình để điều hướng chùm sóng (trung bình tạo ra SER có ích hơn), và do đó, việc giảm âm lượng loa của thiết bị âm thanh 805 sẽ ít ảnh hưởng hơn so với việc giảm âm lượng loa của âm thanh thiết bị 802. Theo một số ví dụ như vậy, QUYẾT ĐỊNH tối ưu là giảm âm lượng loa 806.

Ví dụ khác sẽ được mô tả dựa vào hình vẽ trên Fig.9. Trong trường hợp này, có lẽ sáng chế xem xét sự khác biệt đáng kể nhất của SER cơ bản sẽ xuất hiện trên hai thiết bị: cặp tai nghe, và loa thông minh.

Fig.9 minh họa tình huống trong đó thiết bị SER rất cao rất gần người dùng. Trong hình vẽ trên Fig.9, người dùng 101 đeo tai nghe 902 và nói (phát ra âm thanh 102) theo cách được thu bằng cả micrô 903 trên tai nghe 902 và cả micrô của thiết bị loa thông minh 904. Trong trường hợp này, thiết bị loa thông minh 904 cũng có thể tạo ra một số âm thanh để khớp với tai nghe (ví dụ, kết xuất gần/xa cho âm thanh đắm chìm). Rõ ràng là tai nghe 902 là thiết bị đầu ra gần nhất với người dùng 101, tuy nhiên gần như không có đường dẫn tiếng vọng từ tai nghe đến micrô gần nhất 903, do đó SER cho thiết bị này là rất cao, và ảnh hưởng rất lớn nếu thiết bị được giảm âm lượng vì tai nghe gần như là toàn bộ hiệu ứng kết xuất cho người nghe. Trong trường hợp này, việc giảm âm lượng loa thông minh 904 sẽ có lợi nhiều hơn, mặc dù chỉ có lợi một chút và không thay đổi để kết xuất toàn bộ (những người nghe khác ở gần nghe thấy âm thanh), có thể không có hành động thực tế nào được quyết định - đó là giảm âm lượng loa, hoặc thay đổi các thông số xử lý âm thanh, để có thể cải thiện SER của bộ thu biến của người dùng theo cách mà sự thay đổi âm thanh được tạo ra trong môi trường âm thanh sẽ tốt hơn - theo một nghĩa nào đó, nó đã hoạt động khá tốt do thiết bị SER vốn có trong tai nghe.

Đối với các thiết bị có nhiều loa và các micrô phân tán, ngoài kích thước nhất định, trong một số trường hợp, sáng chế có thể xem xét một thiết bị âm thanh có nhiều loa và nhiều micrô là chòm điểm của các thiết bị riêng được kết nối chặt chẽ với nhau.

Trong trường hợp này, quyết định giảm âm lượng có thể áp dụng cho từng loa. Do đó, theo một số phương án thực hiện, bộ quản lý phiên âm thanh có thể xem loại thiết bị âm thanh này là tập hợp các micrô và loa riêng, còn trong các ví dụ khác, bộ quản lý phiên âm thanh có thể xem loại thiết bị âm thanh này là một thiết bị có các mảng loa và micrô hỗn hợp. Ngoài ra, có thể thấy rằng có tính đối ngẫu giữa việc coi các loa trên một thiết bị là các thiết bị riêng và ý tưởng trong một thiết bị âm thanh nhiều loa, phương pháp kết xuất là điều khiển không gian, việc này nhất thiết phải tạo ra sự thay đổi khác biệt đối với các đầu ra của loa trên một thiết bị âm thanh.

Đối với các hiệu ứng phụ của (các) thiết bị âm thanh gần nhất tránh độ nhạy của hình ảnh không gian từ các thiết bị âm thanh gần người nghe đang di chuyển, trong nhiều trường hợp, ngay cả khi đặt loa gần người nghe đang di chuyển, có thể không có nghĩa để tái tạo các đối tượng âm thanh cụ thể hoặc kết xuất âm thanh ra từ (các) loa gần nhất. Việc này đơn giản là liên quan đến thực tế rằng âm lượng của đường dẫn âm thanh trực tiếp thay đổi trực tiếp là $1/r^2$, trong đó r là khoảng cách âm thanh lan truyền, và khi loa trở nên gần với người nghe bất kỳ ($r \rightarrow 0$) thì sự ổn định của mức âm thanh được tái tạo bằng loa này, so với việc trộn toàn bộ, trở nên kém.

Trong một số trường hợp như vậy, có thể có lợi nếu thực hiện phương án trong đó (ví dụ):

- NGŨ CẢNH là khu vực nghe chung nào đó (ví dụ, ghế dài gần TV), trong đó nó được cho là luôn hữu ích để có thể nghe được âm thanh của chương trình mà người nào đó đang xem trên TV;
- QUYẾT ĐỊNH: đối với thiết bị có loa trên bàn cà phê trong khu vực nghe chung (ví dụ, gần ghế dài), đặt $f_n = 1$; và
- KẾT XUẤT: Thiết bị được tắt, và năng lượng được kết xuất ở nơi khác.

Ảnh hưởng của sự thay đổi xử lý âm thanh này là giúp người ngồi trên ghế nghe tốt hơn. Nếu bàn cà phê được đặt ở một bên của ghế dài, thì phương pháp này sẽ tránh sự nhạy cảm của người nghe khi ở gần thiết bị âm thanh này. Trong khi thiết bị âm thanh này, trong một số trường hợp, có thể có vị trí lý tưởng, ví dụ, kênh âm thanh vòm, thực tế là có thể có chênh lệch mức 20 dB hoặc hơn trên ghế dài so với loa này tức là trừ khi biết chính xác vị trí của người nghe/người nói, có thể là ý tưởng tốt để giảm âm lượng hoặc tắt thiết bị gần nhất này.

Fig.10 là lưu đồ mô tả một ví dụ của phương pháp mà có thể được thực hiện bằng thiết bị như được thể hiện trên Fig.2A. Các khối của phương pháp 1000, giống như các phương pháp khác được mô tả ở đây, không nhất thiết phải được thực hiện theo thứ tự được biểu thị. Hơn nữa, các phương pháp như vậy có thể bao gồm nhiều khối hơn hoặc ít khối hơn so với được thể hiện và/hoặc mô tả. Trong phương án thực hiện này, phương pháp 1000 bao gồm bước ước lượng vị trí của người dùng trong môi trường.

Trong ví dụ này, khối 1005 bao gồm bước nhận tín hiệu đầu ra từ mỗi micrô trong số nhiều micrô trong môi trường. Trong trường hợp này, mỗi trong số nhiều micrô nằm ở vị trí micrô của môi trường. Theo ví dụ này, các tín hiệu đầu ra tương ứng với tiếng nói hiện tại của người dùng. Theo một số ví dụ, tiếng nói hiện tại có thể là, hoặc có thể bao gồm, tiếng nói từ đánh thức. Ví dụ, khối 1005 có thể liên quan đến hệ thống điều khiển (như hệ thống điều khiển 120 trên Fig.2A) nhận tín hiệu đầu ra từ mỗi micrô trong số nhiều micrô trong môi trường qua hệ thống giao diện (như hệ thống giao diện 205 trên Fig.2A).

Theo một số ví dụ, ít nhất một số trong số các micrô trong môi trường có thể cung cấp các tín hiệu đầu ra mà không đồng bộ với các tín hiệu đầu ra do một hoặc nhiều micrô khác cung cấp. Ví dụ, micrô thứ nhất trong số nhiều micrô có thể lấy mẫu dữ liệu âm thanh theo xung nhịp lấy mẫu thứ nhất và micrô thứ hai trong số nhiều micrô có thể lấy mẫu dữ liệu âm thanh theo xung nhịp lấy mẫu thứ hai. Trong một số trường hợp, ít nhất một trong số các micrô trong môi trường có thể được bao gồm, hoặc được tạo cấu hình để truyền thông với thiết bị âm thanh thông minh.

Theo ví dụ này, khối 1010 bao gồm bước xác định nhiều đặc tính âm thanh hiện tại từ các tín hiệu đầu ra của mỗi micrô. Trong ví dụ này, “các đặc tính âm thanh hiện tại” là các đặc tính âm thanh suy ra từ “tiếng nói hiện tại” của khối 1005. Theo một số phương án thực hiện, khối 1010 có thể bao gồm bước nhận nhiều đặc tính âm thanh hiện tại từ một hoặc nhiều thiết bị khác. Ví dụ, khối 1010 có thể bao gồm bước nhận ít nhất một số đặc tính âm thanh hiện tại từ một hoặc nhiều bộ phát hiện từ đánh thức được áp dụng bằng một hoặc nhiều thiết bị khác. Ngoài ra, hoặc cách khác, theo một số phương án thực hiện, khối 1010 có thể bao gồm bước xác định nhiều đặc tính âm thanh hiện tại từ các tín hiệu đầu ra.

Cho dù các đặc tính âm thanh được xác định bằng một thiết bị hay nhiều thiết bị, các đặc tính âm thanh này có thể được xác định là không đồng bộ. Nếu các đặc tính âm thanh này được xác định bằng nhiều thiết bị, thì các đặc tính âm thanh nói chung sẽ được xác định theo cách không đồng bộ trừ khi các thiết bị được tạo cấu hình để phối hợp quy trình xác định các đặc tính âm thanh. Nếu các đặc tính âm thanh được xác định bằng một thiết bị, theo một số phương án thực hiện, các đặc tính âm thanh vẫn có thể được xác định theo cách không đồng bộ vì một thiết bị có thể nhận các tín hiệu đầu ra của mỗi micrô tại các thời điểm khác nhau. Theo một số ví dụ, các đặc tính âm thanh có thể được xác định theo cách không đồng bộ vì ít nhất một số micrô trong môi trường có thể cung cấp các tín hiệu đầu ra không đồng bộ so với các tín hiệu đầu ra do một hoặc nhiều micrô khác cung cấp.

Theo một số ví dụ, các đặc tính âm thanh có thể bao gồm metric độ tin cậy từ đánh thức, metric khoảng thời gian từ đánh thức, và/hoặc ít nhất một metric mức nhận được. Metric mức nhận được có thể biểu thị mức nhận được của âm thanh do micrô phát hiện và có thể tương ứng với mức tín hiệu đầu ra của micrô.

Cách khác hoặc ngoài ra, các đặc tính âm thanh có thể bao gồm một hoặc nhiều đặc điểm sau:

- Entropy trạng thái trung bình (độ tinh khiết) cho mỗi trạng thái từ đánh thức theo đồng chỉnh 1-best (Viterbi) với mô hình âm thanh.
- Hao tổn CTC (Hao tổn phân loại theo thời gian kết nối) so với các mô hình âm thanh của bộ phát hiện từ đánh thức.
- Bộ phát hiện từ đánh thức có thể được huấn luyện để cung cấp ước lượng khoảng cách của người nói từ micrô và/hoặc ước lượng RT60 ngoài độ tin cậy của từ đánh thức. Việc ước lượng khoảng cách và/hoặc ước lượng RT60 có thể là các đặc tính âm thanh.
- Thay vì, hoặc ngoài ra, mức/công suất nhận được bằng rộng tại micrô, đặc tính âm thanh có thể là mức nhận được trong một số băng tần số cách nhau log/Mel/Bark. Các băng tần số này có thể thay đổi tùy theo phương án thực hiện cụ thể (ví dụ, 2 băng tần, 5 băng tần, 20 băng tần, 50 băng tần, băng tần 1 quãng tám hoặc băng tần 1/3 quãng tám).

- Dạng biểu diễn Cepstral của thông tin phổ trong điểm trước đó, được tính toán bằng cách lấy logarit của lũy thừa băng tần cho DCT (biến đổi cosin rời rạc - discrete cosine transform).

- Lấy lũy thừa băng tần trong các băng tần có trọng số cho lời nói con người. Ví dụ, các đặc tính âm thanh có thể chỉ được dựa trên một băng tần cụ thể (ví dụ, từ 400 Hz đến 1,5 kHz). Trong ví dụ này, các tần số cao hơn và thấp hơn có thể được bỏ qua.

- Độ tin cậy của bộ phát hiện hoạt động giọng nói trên mỗi băng tần hoặc trên mỗi bin.

- Các đặc tính âm thanh có thể được dựa, ít nhất một phần, vào ước lượng tiếng ồn dài hạn để bỏ qua các micrô có tỷ lệ tín hiệu trên tiếng ồn kém.

- Độ nhọn là số đo của “đỉnh” lời nói. Độ nhọn có thể là chỉ báo làm nhòe bằng đuôi âm dội dài.

- Thời điểm bắt đầu từ đánh thức được ước lượng. Có thể kỳ vọng lúc bắt đầu và khoảng thời gian bằng nhau trong một khung hoặc trên tất cả các micrô. Khoảng lệch có thể đưa ra manh mối về ước lượng không đáng tin cậy. Việc này giả định mức đồng bộ nào đó - không nhất thiết đối với mẫu - nhưng, ví dụ, với khung vài chục mili giây.

Theo ví dụ này, khối 1015 bao gồm bước áp dụng bộ phân loại cho nhiều đặc tính âm thanh hiện thời. Theo một số ví dụ như vậy, bước áp dụng bộ phân loại có thể bao gồm bước áp dụng mô hình được huấn luyện trên các đặc tính âm thanh được xác định từ trước thu được từ nhiều tiếng nói trước đó của người dùng trong nhiều vùng của người dùng trong môi trường. Một số ví dụ khác nhau được cung cấp ở đây.

Theo một số ví dụ, vùng của người dùng có thể bao gồm khu vực bồn rửa, khu vực chuẩn bị thực phẩm, khu vực tủ lạnh, khu vực ăn uống, khu vực ghế dài, khu vực xem tivi, khu vực phòng ngủ, và/hoặc khu vực cửa ra vào. Theo một số ví dụ, một hoặc nhiều vùng người dùng có thể là vùng người dùng định trước. Theo một số ví dụ như vậy, một hoặc nhiều vùng người dùng định trước có thể là lựa chọn được bởi người dùng trong quy trình huấn luyện.

Theo một số phương án thực hiện, bước áp dụng bộ phân loại có thể bao gồm bước áp dụng Mô hình hỗn hợp Gauss được huấn luyện trên các tiếng nói trước đó. Theo một số phương án thực hiện như vậy, bước áp dụng bộ phân loại có thể bao gồm bước áp dụng Mô hình hỗn hợp Gauss được huấn luyện trên một hoặc nhiều độ tin cậy từ

đánh thức chuẩn hóa, mức nhận được trung bình chuẩn hóa, hoặc mức nhận được tối đa của các tiếng nói trước đó. Tuy nhiên, trong các phương án thực hiện khác, bước áp dụng bộ phân loại có thể được dựa trên mô hình khác, như một trong số các mô hình khác được mô tả ở đây. Trong một số trường hợp, mô hình có thể được huấn luyện bằng cách sử dụng dữ liệu huấn luyện được gắn nhãn với các vùng người dùng. Tuy nhiên, theo một số ví dụ, bước áp dụng bộ phân loại bao gồm bước áp dụng mô hình được huấn luyện bằng cách sử dụng dữ liệu huấn luyện không có nhãn tức là không được gắn nhãn với các vùng người dùng.

Theo một số ví dụ, các tiếng nói trước đó có thể đã, hoặc có thể đã bao gồm, các tiếng nói từ đánh thức. Theo một số ví dụ như vậy, các tiếng nói trước đây và tiếng nói hiện tại có thể là tiếng nói của cùng một từ đánh thức.

Trong ví dụ này, khối 1020 bao gồm bước xác định, dựa ít nhất một phần vào đầu ra từ bộ phân loại, ước lượng về vùng người dùng trong đó người dùng hiện đang ở. Theo một số ví dụ như vậy, ước lượng có thể được xác định mà không cần tham chiếu đến vị trí hình học của nhiều micrô. Ví dụ, ước lượng có thể được xác định mà không cần tham chiếu đến các tọa độ của các micrô riêng. Theo một số ví dụ, ước lượng có thể được xác định mà không cần ước lượng vị trí hình học của người dùng.

Một số phương án thực hiện của phương pháp 1000 có thể bao gồm bước lựa chọn ít nhất một loa theo vùng người dùng được ước lượng. Một số phương án thực hiện như vậy có thể bao gồm bước điều khiển ít nhất một loa đã chọn để cung cấp âm thanh cho vùng người dùng được ước lượng. Cách khác hoặc ngoài ra, một số phương án thực hiện của phương pháp 1000 có thể bao gồm bước lựa chọn ít nhất một micrô theo vùng người dùng được ước lượng. Một số phương án thực hiện như vậy có thể bao gồm bước cung cấp tín hiệu xuất ra bởi ít nhất một micrô đã chọn cho thiết bị âm thanh thông minh.

Fig.11 là sơ đồ khối của các phần tử của một ví dụ theo một phương án được tạo cấu hình để áp dụng bộ phân loại vùng. Theo ví dụ này, hệ thống 1100 bao gồm nhiều loa 1104 được phân bố trong ít nhất một phần của môi trường (ví dụ, môi trường như được minh họa trên Fig.1A hoặc Fig.1B). Trong ví dụ này, hệ thống 1100 bao gồm bộ kết xuất loa đa kênh 1101. Theo phương án thực hiện này, đầu ra của bộ kết xuất loa đa kênh 1101 đóng vai trò là cả tín hiệu điều khiển loa (nguồn cấp loa để điều khiển các loa 1104) và các mẫu tiếng vọng. Trong phương án thực hiện này, các mẫu tiếng vọng

được cung cấp cho hệ thống phụ quản lý tiếng vọng 1103 qua nhiều kênh tham chiếu loa 1102, mà bao gồm ít nhất một số tín hiệu cấp cho loa xuất ra từ bộ kết xuất 1101.

Trong phương án thực hiện này, hệ thống 1100 bao gồm nhiều hệ thống phụ quản lý tiếng vọng 1103. Theo ví dụ này, các hệ thống phụ quản lý tiếng vọng 1103 được tạo cấu hình để thực hiện một hoặc nhiều quy trình khử tiếng vọng và/hoặc một hoặc nhiều quy trình triệt tiếng vọng. Trong ví dụ này, mỗi hệ thống phụ quản lý tiếng vọng 1103 cung cấp đầu ra quản lý tiếng vọng 1103A tương ứng với một trong số các bộ phát hiện từ đánh thức 1106. Đầu ra quản lý tiếng vọng 1103A đã làm suy yếu tiếng vọng so với đầu vào với hệ thống phụ liên quan đến trong số các hệ thống phụ quản lý tiếng vọng 1103.

Theo phương án thực hiện này, hệ thống 1100 bao gồm N micrô 1105 (N là số nguyên) được phân phối trong ít nhất một phần của môi trường (ví dụ, môi trường được minh họa trên Fig.1A hoặc Fig.1B). Các micrô có thể bao gồm micrô mảng và/hoặc micrô điểm. Ví dụ, một hoặc nhiều thiết bị âm thanh thông minh đặt trong môi trường có thể bao gồm mảng micrô. Trong ví dụ này, đầu ra của micrô 1105 được cung cấp làm đầu vào cho hệ thống phụ quản lý tiếng vọng 1103. Theo phương án thực hiện này, mỗi trong số các hệ thống phụ quản lý tiếng vọng 1103 thu nhận đầu ra của một micrô 1105 riêng hoặc nhóm hoặc tập hợp con micrô 1105 riêng).

Trong ví dụ này, hệ thống 1100 bao gồm nhiều bộ phát hiện từ đánh thức 1106. Theo ví dụ này, mỗi bộ phát hiện từ đánh thức 1106 nhận đầu ra âm thanh từ một trong số các hệ thống phụ quản lý tiếng vọng 1103 và xuất ra nhiều đặc tính âm thanh 1106A. Các đặc tính âm thanh 1106A xuất ra từ mỗi hệ thống phụ quản lý tiếng vọng 1103 có thể bao gồm (nhưng không giới hạn ở): độ tin cậy của từ đánh thức, khoảng thời gian của từ đánh thức và các số đo mức nhận được. Mặc dù ba mũi tên, mô tả ba đặc tính âm thanh 1106A, được thể hiện là được xuất ra từ mỗi hệ thống phụ quản lý tiếng vọng 1103, nhiều hơn hoặc ít hơn các đặc tính âm thanh 1106A có thể được xuất ra trong các phương án thực hiện khác. Hơn nữa, mặc dù ba mũi tên này chạm vào bộ phân loại 1107 dọc theo nhiều hoặc ít đường thẳng đứng, nhưng việc này không biểu thị rằng bộ phân loại 1107 nhất thiết nhận các đặc tính âm thanh 1106A từ tất cả các bộ phát hiện từ đánh thức 1106 cùng một thời điểm. Như đã lưu ý ở phần khác ở đây, các đặc tính âm thanh

1106A có thể, trong một số trường hợp, được xác định và/hoặc được cung cấp cho bộ phân loại theo cách không đồng bộ.

Theo phương án thực hiện này, hệ thống 1100 bao gồm bộ phân loại vùng 1107, cũng có thể được gọi là bộ phân loại 1107. Trong ví dụ này, bộ phân loại nhận nhiều đặc tính 1106A từ nhiều bộ phát hiện từ đánh thức 1106 cho nhiều (ví dụ, tất cả các) micrô 1105 trong môi trường. Theo ví dụ này, đầu ra 1108 của bộ phân loại vùng 1107 tương ứng với ước lượng về vùng người dùng mà người dùng hiện đang ở. Theo một số ví dụ như vậy, đầu ra 1108 có thể tương ứng với một hoặc nhiều xác suất hậu nghiệm. Ước lượng về vùng người dùng mà người dùng hiện đang ở có thể là, hoặc có thể tương ứng với, xác suất *hậu nghiệm* tối đa theo thống kê Bayesian.

Tiếp theo, sáng chế mô tả các phương án thực hiện làm ví dụ của bộ phân loại, mà có thể trong một số ví dụ tương ứng với bộ phân loại vùng 1107 trên Fig.11. Gọi $x_i(n)$ là tín hiệu micrô thứ i , $i = \{1 \dots N\}$, tại thời điểm rời rạc n (ví dụ, tín hiệu micrô $x_i(n)$ là đầu ra của N micrô 1105). Xử lý N tín hiệu $x_i(n)$ trong hệ thống phụ quản lý tiếng vọng 1103 tạo ra tín hiệu micrô 'sạch' $e_i(n)$, trong đó $i = \{1 \dots N\}$, mỗi tín hiệu tại một thời điểm riêng n . Tín hiệu sạch $e_i(n)$, được gọi là 1103A trên Fig.11, được cấp cho bộ phát hiện từ đánh thức 1106 trong ví dụ này. Ở đây, mỗi bộ phát hiện từ đánh thức 1106 cung cấp vector đặc tính $\mathbf{w}_i(j)$, được gọi là 1106A trên Fig.11, trong đó $j = \{1 \dots J\}$ là chỉ số tương ứng với tiếng nói từ đánh thức thứ j . Trong ví dụ này, bộ phân loại 1107 lấy đầu vào là tập hợp đặc tính tổng hợp $\mathbf{W}(j) = [\mathbf{w}_1^T(j) \dots \mathbf{w}_N^T(j)]^T$.

Theo một số phương án thực hiện, tập hợp các nhãn vùng C_k , đối với $k = \{1 \dots K\}$, có thể tương ứng với số vùng người dùng khác nhau, K , trong môi trường. Ví dụ, các vùng người dùng có thể bao gồm khu vực ghế dài, khu vực bếp, khu vực ghế đọc sách, v.v. Một số ví dụ có thể xác định nhiều hơn một khu vực trong bếp hoặc phòng khác. Ví dụ, khu vực bếp có thể bao gồm khu bồn rửa, khu chuẩn bị thức ăn, khu tủ lạnh và khu ăn uống. Tương tự, khu vực phòng khách có thể bao gồm khu vực ghế dài, khu vực xem tivi, khu vực ghế đọc sách, một hoặc nhiều khu vực cửa ra vào, v.v. Các nhãn vùng cho các vùng này có thể lựa chọn được bởi người dùng, ví dụ, trong pha huấn luyện.

Theo một số phương án thực hiện, bộ phân loại 1107 ước lượng các xác suất hậu nghiệm $p(C_k|\mathbf{W}(j))$ của tập hợp đặc tính $\mathbf{W}(j)$, ví dụ, bằng cách sử dụng bộ phân loại

Bayesian. Các xác suất $p(C_k|\mathbf{W}(j))$ biểu thị xác suất (đối với tiếng nói thứ j và vùng thứ k , cho mỗi vùng trong số các vùng C_k , và mỗi trong số các tiếng nói) mà người dùng ở trong mỗi vùng C , và là một ví dụ của đầu ra 1108 của bộ phân loại 1107.

Theo một số ví dụ, dữ liệu huấn luyện có thể được tập hợp (ví dụ, cho từng vùng người dùng) bằng cách nhắc người dùng lựa chọn hoặc xác định vùng, ví dụ, vùng ghế dài. Quy trình huấn luyện có thể bao gồm bước nhắc người dùng tạo ra tiếng nói huấn luyện, như từ đánh thức, trong vùng lân cận của vùng được chọn hoặc được xác định. Theo một ví dụ về khu vực ghế dài, quy trình huấn luyện có thể bao gồm bước nhắc người dùng tạo ra tiếng nói huấn luyện ở trung tâm và các cạnh ở xa của ghế dài. Quy trình huấn luyện có thể bao gồm bước nhắc người dùng lặp lại tiếng nói huấn luyện nhiều lần tại mỗi vị trí trong vùng người dùng. Sau đó, người dùng có thể được nhắc để chuyển sang vùng người dùng khác và tiếp tục đến khi tất cả các vùng người dùng chỉ định được bao phủ.

Fig.12 là lưu đồ mô tả một ví dụ về phương pháp có thể được thực hiện bằng thiết bị như thiết bị 200 trên Fig.2A. Các khối của phương pháp 1200, tương tự các phương pháp khác được mô tả ở đây, không cần thiết được thực hiện theo thứ tự biểu thị. Hơn nữa, các phương pháp như vậy có thể bao gồm nhiều khối hoặc ít khối hơn so với phương pháp được thể hiện và/hoặc mô tả. Trong phương án thực hiện này, phương pháp 1200 bao gồm bước huấn luyện bộ phân loại để ước lượng vị trí của người dùng trong môi trường.

Trong ví dụ này, khối 1205 bao gồm bước nhắc người dùng tạo ra ít nhất một tiếng nói huấn luyện ở mỗi trong số nhiều vị trí trong vùng người dùng thứ nhất của môi trường. (Các) tiếng nói huấn luyện, theo một số ví dụ, có thể là một hoặc nhiều trường hợp của tiếng nói từ đánh thức. Theo một số phương án thực hiện, vùng người dùng thứ nhất có thể là vùng người dùng bất kỳ do người dùng lựa chọn và/hoặc xác định. Trong một số trường hợp, hệ thống điều khiển có thể tạo ra nhãn vùng tương ứng (ví dụ, trường hợp tương ứng của một trong số các nhãn vùng C_k mô tả ở trên) và có thể kết hợp nhãn vùng với dữ liệu huấn luyện thu được cho vùng người dùng thứ nhất.

Hệ thống nhắc tự động có thể được dùng để tập hợp dữ liệu huấn luyện này. Như lưu ý ở trên, hệ thống giao diện 205 của thiết bị 200 có thể bao gồm một hoặc nhiều thiết bị để áp dụng giao diện người dùng, như một hoặc nhiều micro, một hoặc nhiều

loa, hệ thống hiển thị, hệ thống cảm biến chạm, và/hoặc hệ thống cảm biến cử chỉ. Ví dụ, thiết bị 200 có thể cung cấp cho người dùng lời nhắc sau trên màn hình của hệ thống hiển thị hoặc nghe chúng được thông báo qua một hoặc nhiều loa trong quy trình huấn luyện:

- “Đi đến ghế dài.”
- “Nói mười lần từ đánh thức trong khi di chuyển đầu của bạn.”
- “Đi đến vị trí giữa ghế dài và ghế đọc sách và nói mười lần từ đánh thức.”
- “Đứng trong bếp như đang nấu ăn và nói mười lần từ đánh thức.”

Trong ví dụ này, khối 1210 bao gồm bước nhận các tín hiệu đầu ra thứ nhất từ mỗi trong số nhiều micrô trong môi trường. Theo một số ví dụ, khối 1210 có thể bao gồm bước nhận tín hiệu đầu ra thứ nhất từ tất cả các micrô đang hoạt động trong môi trường, còn trong các ví dụ khác, khối 1210 có thể bao gồm bước nhận tín hiệu đầu ra thứ nhất từ tập hợp con của tất cả các micrô đang hoạt động trong môi trường. Theo một số ví dụ, ít nhất một số trong số các micrô trong môi trường có thể cung cấp các tín hiệu đầu ra không đồng bộ với các tín hiệu đầu ra do một hoặc nhiều micrô khác cung cấp. Ví dụ, micrô thứ nhất trong số nhiều micrô có thể lấy mẫu dữ liệu âm thanh theo xung nhịp lấy mẫu thứ nhất và micrô thứ hai trong số nhiều micrô có thể lấy mẫu dữ liệu âm thanh theo xung nhịp lấy mẫu thứ hai.

Trong ví dụ này, mỗi micrô trong số nhiều micrô nằm trong vị trí micrô của môi trường. Trong ví dụ này, các tín hiệu đầu ra thứ nhất tương ứng với các trường hợp của các tiếng nói huấn luyện được phát hiện nhận được từ vùng người dùng thứ nhất. Vì khối 1205 bao gồm bước nhắc người dùng tạo ra ít nhất một tiếng nói huấn luyện trong mỗi trong số nhiều vị trí trong vùng người dùng thứ nhất của môi trường, trong ví dụ này, thuật ngữ “tín hiệu đầu ra thứ nhất” để chỉ tập hợp gồm tất cả các tín hiệu đầu ra tương ứng với các tiếng nói huấn luyện cho vùng người dùng thứ nhất. Trong các ví dụ khác, thuật ngữ “tín hiệu đầu ra thứ nhất” có thể chỉ tập hợp con gồm tất cả các tín hiệu đầu ra tương ứng với các tiếng nói huấn luyện cho vùng người dùng thứ nhất.

Theo ví dụ này, khối 1215 bao gồm bước xác định một hoặc nhiều đặc tính âm thanh thứ nhất từ mỗi tín hiệu đầu ra thứ nhất. Theo một số ví dụ, các đặc tính âm thanh thứ nhất có thể bao gồm metric độ tin cậy từ đánh thức và/hoặc metric mức nhận được. Ví dụ, các đặc tính âm thanh thứ nhất có thể bao gồm metric độ tin cậy từ đánh thức

chuẩn hóa, chỉ báo về mức nhận được trung bình chuẩn hóa và/hoặc chỉ báo về mức nhận được tối đa.

Như lưu ý ở trên, vì khối 1205 bao gồm bước nhắc người dùng tạo ra ít nhất một tiếng nói huấn luyện ở mỗi trong số nhiều vị trí trong vùng người dùng thứ nhất của môi trường, trong ví dụ này, thuật ngữ “tín hiệu đầu ra thứ nhất” đề cập đến tập hợp gồm tất cả tín hiệu đầu ra tương ứng với tiếng nói huấn luyện cho vùng người dùng thứ nhất. Do đó, trong ví dụ này, thuật ngữ “các đặc tính âm thanh thứ nhất” đề cập đến tập hợp các đặc tính âm thanh suy ra từ tập hợp gồm tất cả các tín hiệu đầu ra tương ứng với tiếng nói huấn luyện cho vùng người dùng thứ nhất. Do đó, trong ví dụ này, tập hợp các đặc tính âm thanh thứ nhất ít nhất là lớn bằng tập hợp của các tín hiệu đầu ra thứ nhất. Ví dụ, nếu hai đặc tính âm thanh được xác định từ mỗi trong số các tín hiệu đầu ra, thì tập hợp các đặc tính âm thanh thứ nhất sẽ lớn gấp đôi so với tập hợp các tín hiệu đầu ra thứ nhất.

Trong ví dụ này, khối 1220 bao gồm bước huấn luyện mô hình bộ phân loại để tạo ra sự tương quan giữa vùng người dùng thứ nhất và các đặc tính âm thanh thứ nhất. Ví dụ, mô hình bộ phân loại có thể là bất kỳ trong số các mô hình bộ phân loại được mô tả ở đây. Theo phương án thực hiện này, mô hình bộ phân loại được huấn luyện mà không cần dựa vào các vị trí hình học của nhiều micrô. Nói cách khác, trong ví dụ này, dữ liệu liên quan đến vị trí hình học của nhiều micrô (ví dụ, dữ liệu tọa độ micrô) không được cung cấp cho mô hình bộ phân loại trong quy trình huấn luyện.

Fig.13 là lưu đồ mô tả một ví dụ khác về phương pháp mà có thể được thực hiện bằng thiết bị như thiết bị 200 trên Fig.2A. Các khối của phương pháp 1300, tương tự các phương pháp khác được mô tả ở đây, không cần thiết được thực hiện theo thứ tự biểu thị. Ví dụ, theo một số phương án thực hiện, ít nhất một phần của quy trình xác định đặc tính âm thanh của khối 1325 có thể được thực hiện trước khối 1315 hoặc khối 1320. Hơn nữa, các phương pháp như vậy có thể bao gồm nhiều khối hoặc ít khối hơn so với phương pháp được thể hiện và/hoặc mô tả. Trong phương án thực hiện này, phương pháp 1300 bao gồm bước huấn luyện bộ phân loại để ước lượng vị trí của người dùng trong môi trường. Phương pháp 1300 cung cấp một ví dụ về phương pháp mở rộng 1200 cho nhiều vùng người dùng của môi trường.

Theo ví dụ này, khối 1305 bao gồm bước nhắc người dùng tạo ra ít nhất một tiếng nói huấn luyện ở vị trí trong vùng người dùng của môi trường. Trong một số trường hợp, khối 1305 có thể được thực hiện theo cách được mô tả ở trên dựa vào khối 1205 trên Fig.12, ngoại trừ khối 1305 gắn với một vị trí trong vùng người dùng. (Các) tiếng nói huấn luyện có thể, theo một số ví dụ, là một hoặc nhiều trường hợp của tiếng nói từ đánh thức. Theo một số phương án thực hiện, vùng người dùng có thể là vùng người dùng bất kỳ do người dùng lựa chọn và/hoặc xác định. Trong một số trường hợp, hệ thống điều khiển có thể tạo ra nhãn vùng tương ứng (ví dụ, trường hợp tương ứng của một trong các nhãn vùng C_k được mô tả ở trên) và có thể kết hợp nhãn vùng với dữ liệu huấn luyện thu được cho vùng người dùng.

Theo ví dụ này, khối 1310 được thực hiện cơ bản như mô tả ở trên dựa vào khối 1210 trên Fig.12. Tuy nhiên, theo ví dụ này, quy trình của khối 1310 được tổng quát hóa cho vùng người dùng bất kỳ, không nhất thiết là vùng người dùng thứ nhất mà thu được dữ liệu huấn luyện. Do đó, các tín hiệu đầu ra nhận được trong khối 1310 là “tín hiệu đầu ra từ mỗi trong số nhiều micrô trong môi trường, mỗi trong số nhiều micrô nằm ở vị trí micrô của môi trường, tín hiệu đầu ra tương ứng với các trường hợp các tiếng nói huấn luyện phát hiện được nhận được từ vùng người dùng.” Trong ví dụ này, thuật ngữ “tín hiệu đầu ra” chỉ tập hợp tất cả các tín hiệu đầu ra tương ứng với một hoặc nhiều tiếng nói huấn luyện trong một vị trí của vùng người dùng. Trong các ví dụ khác, thuật ngữ “tín hiệu đầu ra” có thể chỉ tập hợp con của tất cả các tín hiệu đầu ra tương ứng với một hoặc nhiều tiếng nói huấn luyện trong một vị trí của vùng người dùng.

Theo ví dụ này, khối 1315 bao gồm bước xác định xem có thu được đủ dữ liệu huấn luyện cho vùng người dùng hiện tại hay không. Theo một số ví dụ như vậy, khối 1315 có thể bao gồm bước xác định xem các tín hiệu đầu ra có tương ứng với số tiếng nói huấn luyện ngưỡng thu được cho vùng người dùng hiện tại hay không. Cách khác, hoặc ngoài ra, khối 1315 có thể bao gồm bước xác định xem các tín hiệu đầu ra có tương ứng với các tiếng nói huấn luyện ở số lượng vị trí ngưỡng trong vùng người dùng hiện tại đã được thu hay không. Nếu không, phương pháp 1300 trở lại khối 1305 trong ví dụ này và người dùng được nhắc tạo ra ít nhất một tiếng nói bổ sung tại một vị trí trong cùng vùng người dùng.

Tuy nhiên, nếu trong khối 1315 xác định rằng đã có đủ dữ liệu huấn luyện cho vùng người dùng hiện tại, thì trong ví dụ này, quy trình tiếp tục đến khối 1320. Theo ví dụ này, khối 1320 bao gồm bước xác định xem có lấy dữ liệu huấn luyện cho các vùng người dùng bổ sung hay không. Theo một số ví dụ, khối 1320 có thể bao gồm bước xác định xem dữ liệu huấn luyện đã được lấy cho từng vùng người dùng mà người dùng đã được xác định từ trước chưa. Trong các ví dụ khác, khối 1320 có thể bao gồm bước xác định xem dữ liệu huấn luyện đã được lấy cho số lượng vùng người dùng tối thiểu chưa. Số lượng tối thiểu có thể được người dùng lựa chọn. Trong các ví dụ khác, số lượng tối thiểu có thể là số lượng tối thiểu được khuyến nghị cho mỗi môi trường, số lượng tối thiểu được khuyến nghị cho mỗi phòng của môi trường, v.v.

Nếu trong khối 1320 xác định rằng dữ liệu huấn luyện cần được lấy cho các vùng người dùng bổ sung, thì trong ví dụ này, quy trình tiếp tục đến khối 1322, khối này bao gồm bước nhắc người dùng đi đến vùng người dùng khác của môi trường. Theo một số ví dụ, vùng người dùng tiếp theo có thể lựa chọn được bởi người dùng. Theo ví dụ này, quy trình tiếp tục đến khối 1305 sau lời nhắc ở khối 1322. Theo một số ví dụ như vậy, người dùng có thể được nhắc để xác nhận rằng người dùng đã đến vùng người dùng mới sau lời nhắc của khối 1322. Theo một số ví dụ như vậy, người dùng có thể được yêu cầu xác nhận rằng người dùng đã đến vùng người dùng mới trước khi đưa ra lời nhắc ở khối 1305.

Nếu trong khối 1320 xác định rằng không cần lấy dữ liệu huấn luyện cho các vùng người dùng bổ sung, thì trong ví dụ này, quy trình tiếp tục đến khối 1325. Trong ví dụ này, phương pháp 1300 bao gồm bước thu nhận dữ liệu huấn luyện cho K vùng người dùng. Theo phương án thực hiện này, khối 1325 bao gồm bước xác định các đặc tính âm thanh từ thứ nhất đến thứ G từ các tín hiệu đầu ra từ thứ nhất đến thứ H tương ứng với mỗi vùng người dùng trong số các vùng người dùng từ thứ nhất đến thứ K mà dữ liệu huấn luyện đã được thu nhận. Trong ví dụ này, thuật ngữ “tín hiệu đầu ra thứ nhất” đề cập đến tập hợp tất cả các tín hiệu đầu ra tương ứng với các tiếng nói huấn luyện cho vùng người dùng thứ nhất và thuật ngữ “tín hiệu đầu ra thứ H ” đề cập đến tập hợp tất cả các tín hiệu đầu ra tương ứng với các tiếng nói huấn luyện cho vùng người dùng thứ K . Tương tự, thuật ngữ “đặc tính âm thanh thứ nhất” đề cập đến tập hợp các đặc tính âm thanh được xác định từ các tín hiệu đầu ra thứ nhất và thuật ngữ “đặc tính

âm thanh thứ G'' đề cập đến tập hợp các đặc tính âm thanh được xác định từ các tín hiệu đầu ra thứ H .

Theo các ví dụ này, khối 1330 bao gồm bước huấn luyện mô hình bộ phân loại để tạo ra sự tương quan tương ứng giữa các vùng người dùng từ thứ nhất đến thứ K và các đặc tính âm thanh từ thứ nhất đến thứ K . Ví dụ, mô hình bộ phân loại có thể, là bất kỳ trong số các mô hình bộ phân loại được nêu trong sáng chế này.

Trong ví dụ trên, các vùng người dùng được gắn nhãn (ví dụ, theo trường hợp tương ứng của một trong các nhãn vùng C_k được mô tả ở trên). Tuy nhiên, mô hình có thể được huấn luyện theo vùng người dùng được gắn nhãn hoặc không được gắn nhãn, tùy thuộc vào phương án thực hiện cụ thể. Trong trường hợp được gắn nhãn, mỗi tiếng nói huấn luyện có thể được ghép với nhãn tương ứng với vùng người dùng, ví dụ, như sau:

$$p(C_k | \mathbf{W}(j)) = \begin{cases} 1 & k = k' \\ 0 & \text{nếu không} \end{cases}$$

Việc huấn luyện mô hình bộ phân loại có thể bao gồm bước xác định sự phù hợp nhất cho dữ liệu huấn luyện được gắn nhãn. Không làm mất tính tổng quát, các phương pháp phân loại thích hợp cho mô hình bộ phân loại có thể bao gồm:

- Bộ phân loại của Bayes, ví dụ với sự phân phối trên mỗi lớp được mô tả bằng các kỹ thuật phân phối chuẩn nhiều chiều, Mô hình Hỗn hợp Gauss hiệp phương sai đầy đủ hoặc Mô hình Hỗn hợp Gauss hiệp phương sai-đường chéo;
- Lượng tử hóa vector
- Lân cận gần nhất (giá trị trung bình k);
- Mạng nơron có lớp đầu ra SoftMax với một đầu ra tương ứng với mỗi lớp;
- Máy Vector hỗ trợ (Support Vector Machine – SVM); và/hoặc
- Các kỹ thuật tăng cường, như các Máy tăng cường gradient (Gradient Boosting Machine – GBM)

Theo một ví dụ của phương án thực hiện trường hợp không được gắn nhãn, dữ liệu có thể tự động được chia thành K cụm, trong đó K cũng có thể không được biết. Ví dụ, việc tách tự động không được gắn nhãn có thể được thực hiện bằng cách sử dụng kỹ thuật phân cụm cổ điển, ví dụ, thuật toán k-mean hoặc Mô hình hỗn hợp Gauss.

Để cải thiện tính chắc chắn, việc chuẩn hóa có thể được áp dụng cho việc huấn luyện mô hình bộ phân loại và các tham số mô hình có thể được cập nhật theo thời gian khi các tiếng nói mới được tạo ra.

Tiếp theo, sáng chế mô tả các khía cạnh khác của một phương án.

Tập hợp đặc tính âm thanh làm ví dụ (ví dụ, các đặc tính âm thanh 1106A trên Fig.11) có thể bao gồm khả năng của độ tin cậy từ đánh thức, mức nhận được trung bình trong khoảng thời gian được ước lượng của từ đánh thức tin cậy nhất, và mức nhận được tối đa trong khoảng thời gian của từ đánh thức tin cậy nhất. Các đặc tính có thể được chuẩn hóa liên quan đến các giá trị tối đa của chúng cho mỗi tiếng nói từ đánh thức. Dữ liệu huấn luyện có thể được gắn nhãn và mô hình hỗn hợp Gauss (Gaussian Mixture Model – GMM) hiệp phương sai đầy đủ có thể được huấn luyện để tối đa hóa giá trị kỳ vọng của các nhãn huấn luyện. Vùng được ước lượng có thể là lớp tối đa hóa xác suất hậu nghiệm.

Phần mô tả ở trên của một số phương án mô tả về việc học mô hình vùng âm thanh từ tập hợp dữ liệu huấn luyện được thu thập trong quy trình thu thập nhắc nhở. Trong mô hình đó, thời gian huấn luyện (hoặc chế độ cấu hình) và thời gian chạy (hoặc chế độ thường xuyên) có thể được coi là hai chế độ khác nhau mà hệ thống micro có thể được đặt vào. Phần mở rộng cho sơ đồ này là học trực tuyến, trong đó một số hoặc tất cả mô hình vùng âm thanh được học hoặc được làm thích ứng trực tuyến (ví dụ, tại thời gian chạy hoặc ở chế độ thông thường). Nói cách khác, ngay cả khi bộ phân loại được áp dụng trong quy trình “thời gian chạy” để tạo ra ước lượng về vùng người dùng mà người dùng hiện đang ở (ví dụ, theo phương pháp 1000 trên Fig.10, theo một số phương án thực hiện, quy trình huấn luyện bộ phân loại có thể tiếp tục.

Fig.14 là lưu đồ mô tả một ví dụ khác về phương pháp có thể được thực hiện bằng thiết bị như thiết bị 200 trên Fig.2A. Các khối của phương pháp 1400, tương tự các phương pháp khác được mô tả ở đây, không cần thiết được thực hiện theo thứ tự biểu thị. Hơn nữa, các phương pháp như vậy có thể bao gồm nhiều khối hoặc ít khối hơn so với phương pháp được thể hiện và/hoặc mô tả. Trong phương án thực hiện này, phương pháp 1400 bao gồm bước huấn luyện liên tục bộ phân loại trong quy trình “thời gian chạy” để ước lượng vị trí của người dùng trong môi trường. Phương pháp 1400 là ví dụ về phương pháp được mô tả ở đây là chế độ học trực tuyến.

Trong ví dụ này, khối 1405 của phương pháp 1400 tương ứng với các khối từ 1005 đến 1020 của phương pháp 1000. Ở đây, khối 1405 bao gồm bước tạo ra ước lượng, dựa ít nhất một phần vào đầu ra từ bộ phân loại, về vùng người dùng mà người dùng hiện đang ở. Theo phương án thực hiện này, khối 1410 bao gồm bước thu nhận phản hồi ngầm hoặc rõ ràng về việc ước lượng ở khối 1405. Trong khối 1415, bộ phân loại được cập nhật theo phản hồi nhận được trong khối 1405. Ví dụ, khối 1415 có thể bao gồm một hoặc nhiều phương pháp học tăng cường. Như được đưa ra bằng mũi tên đứt nét từ khối 1415 đến khối 1405, theo một số phương án thực hiện, phương pháp 1400 có thể bao gồm bước trở về khối 1405. Ví dụ, phương pháp 1400 có thể bao gồm bước cung cấp các ước lượng trong tương lai về vùng người dùng mà người dùng ở tại thời điểm tương lai đó, dựa trên việc áp dụng mô hình cập nhật.

Các kỹ thuật rõ ràng để thu được phản hồi có thể bao gồm:

- Hỏi người dùng xem dự đoán có chính xác hay không bằng cách sử dụng giao diện người dùng (user interface – UI) giọng nói. Ví dụ, âm thanh chỉ báo có thể được cung cấp cho người dùng như sau: “Tôi nghĩ rằng bạn đang ở trên ghế dài, vui lòng nói ‘đúng’ hoặc ‘sai’).
- Thông báo cho người dùng rằng các dự đoán không chính xác có thể được hiệu chỉnh bất kỳ lúc nào bằng cách sử dụng UI giọng nói. (ví dụ, âm thanh chỉ báo có thể được cung cấp cho người dùng như sau: “Hiện giờ tôi có thể dự đoán bạn đang ở đâu khi bạn nói với tôi. Nếu tôi dự đoán sai, chỉ cần nói điều gì đó như ‘Amanda, tôi không ở trên ghế dài. Tôi đang ngồi ở ghế đọc sách’).
- Thông báo cho người dùng rằng các dự đoán chính xác có thể được hoàn lại bất kỳ lúc nào bằng cách sử dụng UI giọng nói. (ví dụ, âm thanh chỉ báo có thể được cung cấp cho người dùng như sau: “Hiện giờ tôi có thể dự đoán bạn đang ở đâu khi bạn nói với tôi. Nếu tôi dự đoán đúng, thì bạn có thể giúp cải thiện thêm dự đoán của tôi bằng cách nói điều gì đó như ‘Amanda, đúng vậy. Tôi đang ở trên ghế dài.’).
- Bao gồm các nút vật lý hoặc các phần tử UI khác mà người dùng có thể thao tác để đưa ra phản hồi (ví dụ, nút đồng ý và/hoặc không đồng ý trên thiết bị vật lý hoặc trong ứng dụng điện thoại thông minh).

Mục tiêu của việc dự đoán vùng người dùng mà người dùng đang ở có thể là để thông báo lựa chọn micrô hoặc sơ đồ điều hướng chùm sóng thích ứng để cố gắng thu âm thanh từ vùng âm thanh của người dùng hiệu quả hơn, ví dụ, để nhận dạng lệnh tiếp theo từ đánh thức tốt hơn. Trong các tình huống như vậy, các kỹ thuật ngầm định để thu được phản hồi về chất lượng của dự đoán vùng có thể bao gồm:

- Phạt các dự đoán dẫn đến nhận dạng sai lệnh theo sau từ đánh thức. Proxy mà có thể biểu thị sự nhận dạng sai có thể bao gồm việc người dùng cắt ngắn đáp ứng của trợ lý giọng nói đối với lệnh, ví dụ, nói lệnh ngăn chặn, như “Amanda, dừng lại!”;
- Phạt các dự đoán dẫn đến độ tin cậy thấp mà bộ nhận dạng lời nói đã nhận dạng lệnh thành công. Nhiều hệ thống nhận dạng lời nói tự động có khả năng trả lại mức độ tin cậy với kết quả của chúng có thể được dùng cho mục đích này;
- Phạt các dự đoán dẫn đến việc bộ phát hiện từ đánh thức mật khẩu thứ hai lỗi khi phát hiện lại từ đánh thức có độ tin cậy cao; và/hoặc
- Củng cố các dự đoán dẫn đến sự nhận dạng từ đánh thức có độ tin cậy cao và/hoặc nhận dạng chính xác lệnh của người dùng.

Sau đây là ví dụ về việc bộ phát hiện từ đánh thức mật khẩu thứ hai lỗi khi phát hiện lại từ đánh thức có độ tin cậy cao. Giả sử rằng sau khi thu được tín hiệu đầu ra tương ứng với tiếng nói hiện tại từ các micrô trong môi trường và sau khi xác định các đặc tính âm thanh dựa trên tín hiệu đầu ra (ví dụ, qua nhiều bộ phát hiện từ đánh thức mật khẩu thứ nhất được tạo cấu hình để truyền thông với các micrô), các đặc tính âm thanh sẽ được cung cấp cho bộ phân loại. Nói cách khác, các đặc tính âm thanh được cho là tương ứng với tiếng nói từ đánh thức phát hiện được. Giả sử thêm rằng bộ phân loại xác định rằng người mà có tiếng nói hiện tại nhiều khả năng đang ở vùng 3, vùng này tương ứng với ghế đọc trong ví dụ này. Ví dụ, có thể có một micrô cụ thể hoặc tổ hợp các micrô đã học được biết là tốt nhất để nghe giọng nói của người đó khi người đó trong vùng 3, ví dụ, để gửi cho dịch vụ trợ lý ảo dựa trên đám mây để nhận dạng lệnh bằng giọng nói.

Giả sử thêm rằng sau khi xác định (các) micrô nào sẽ được dùng để nhận dạng lời nói, nhưng trước khi lời nói của người đó thực sự được gửi cho dịch vụ trợ lý ảo, bộ phát hiện từ đánh thức mật khẩu thứ hai hoạt động trên các tín hiệu micrô tương ứng với

lời nói được phát hiện bằng (các) micrô đã chọn cho vùng 3 mà bạn đưa ra để nhận dạng lệnh. Nếu bộ phát hiện từ đánh thức mật khẩu thứ hai đó không phù hợp với nhiều bộ phát hiện từ đánh thức mật khẩu thứ nhất mà từ đánh thức đã thực sự được phát ra thì có thể là do bộ phân loại đã dự đoán sai vùng. Do đó, bộ phân loại cần bị phạt.

Các kỹ thuật để cập nhật *hậu nghiệm* của mô hình lập bản đồ vùng sau khi một hoặc nhiều từ đánh thức đã được nói có thể bao gồm:

- Tối đa sự thích ứng *hậu nghiệm* (Maximum *a posteriori* – MAP) của Mô hình hỗn hợp Gauss (Gaussian Mixture Model – GMM) hoặc mô hình lân cận gần nhất; và/hoặc
- Học củng cố, ví dụ về mạng nơron, ví dụ bằng cách kết hợp nhãn sự thực nền tảng “1 nóng” (nếu dự đoán đúng) hoặc “1 lạnh” (nếu dự đoán sai) thích hợp với đầu ra SoftMax và áp dụng truyền ngược trực tuyến để xác định các trọng số mạng mới.

Một số ví dụ về việc làm thích ứng MAP trong ngữ cảnh này có thể bao gồm bước điều chỉnh phương tiện trong GMM mỗi khi từ đánh thức được nói. Theo cách này, phương tiện có thể trở nên giống với các đặc tính âm thanh quan sát được hơn khi các từ đánh thức tiếp theo được nói. Cách khác, hoặc ngoài ra, các ví dụ như vậy có thể bao gồm bước điều chỉnh phương sai/hiệp phương sai hoặc thông tin trọng số hỗn hợp trong GMM mỗi khi từ đánh thức được nói.

Ví dụ, sơ đồ thích ứng MAP có thể như sau:

$$\mu_{i,\text{new}} = \mu_{i,\text{old}} * \alpha + x * (1 - \alpha)$$

Trong phương trình trên, $\mu_{i,\text{old}}$ biểu diễn giá trị trung bình của Gauss thứ i trong hỗn hợp, α biểu diễn tham số kiểm soát sự thích ứng MAP mạnh mẽ sẽ xảy ra như thế nào (α có thể nằm trong khoảng $[0, 0,999]$) và x biểu diễn vector đặc trưng của tiếng nói từ đánh thức mới. Chỉ số “ i ” tương ứng với phần tử hỗn hợp trả về xác suất *tiên nghiệm* a cao nhất chứa vị trí của loa tại thời điểm có từ đánh thức.

Theo cách khác, mỗi phần tử hỗn hợp có thể được điều chỉnh theo xác suất *tiên nghiệm* a của chúng chứa từ đánh thức, ví dụ, như sau:

$$M_{i,\text{new}} = \mu_{i,\text{old}} * \beta_i * x * (1 - \beta_i)$$

Trong phương trình trên, $\beta_i = \alpha * (1 - P(i))$, trong đó $P(i)$ biểu diễn xác suất *tiên nghiệm* a mà giá trị quan sát x là do phần tử hỗn hợp i .

Theo một ví dụ về học củng cố, có thể có ba vùng người dùng. Giả sử rằng đối với một từ đánh thức cụ thể, mô hình dự đoán các xác suất là $[0,2, 0,1, 0,7]$ cho ba vùng người dùng. Nếu nguồn thông tin thứ hai (ví dụ, bộ phát hiện từ đánh thức mật khẩu thứ hai) xác nhận rằng vùng thứ ba là chính xác, thì nhãn sự thật cơ bản có thể là $[0, 0, 1]$ (“1 nóng”). Bước cập nhật *hậu nghiệm* của mô hình lập bản đồ vùng có thể bao gồm bước truyền ngược lỗi qua mạng nơron, tức là mạng nơron sẽ dự đoán mạnh mẽ hơn vùng 3 nếu đầu vào được thể hiện lần nữa. Ngược lại, theo một ví dụ, nếu nguồn thông tin thứ hai thể hiện rằng vùng 3 là dự đoán không chính xác, thì nhãn sự thật cơ bản có thể là $[0,5, 0,5, 0,0]$. Việc lan truyền ngược lỗi qua mạng nơron sẽ làm cho mô hình ít có khả năng dự đoán vùng 3 hơn nếu được thể hiện cùng một đầu vào trong tương lai.

Việc kết xuất linh hoạt cho phép âm thanh không gian sẽ được kết xuất qua số lượng loa tùy ý được bố trí tùy ý. Theo phương án triển khai rộng của các thiết bị âm thanh, bao gồm nhưng không giới hạn ở các thiết bị âm thanh thông minh (ví dụ, loa thông minh) trong nhà, cần phải hiện thực hóa công nghệ kết xuất linh hoạt cho phép các sản phẩm tiêu dùng thực hiện kết xuất âm thanh linh hoạt, và phát lại của âm thanh đã được kết xuất như vậy.

Một số công nghệ đã được phát triển để thực hiện kết xuất linh hoạt. Các công nghệ này gộp vấn đề kết xuất như một trong những vấn đề tối thiểu hóa hàm chi phí, trong đó hàm chi phí bao gồm hai số hạng: số hạng thứ nhất lập mô hình ấn tượng không gian mong muốn mà bộ kết xuất cố gắng đạt được, và số hạng thứ hai ấn định chi phí để kích hoạt các loa. Đến nay, số hạng thứ hai này đã tập trung vào việc tạo ra giải pháp thừa trong đó chỉ những loa ở gần vị trí không gian mong muốn của âm thanh được kết xuất mới được kích hoạt.

Việc phát lại âm thanh không gian trong môi trường của người tiêu dùng thường được gắn với số lượng loa được chỉ định đặt ở các vị trí được chỉ định: ví dụ, âm thanh vòm 5.1 và 7.1. Trong các trường hợp này, nội dung được tạo riêng cho các loa có liên quan và được mã hóa dưới dạng các kênh rời rạc, một kênh cho mỗi loa (ví dụ, Dolby Digital hoặc Dolby Digital Plus, v.v.). Gần đây, các định dạng âm thanh không gian dựa trên đối tượng, đắm chìm đã được giới thiệu (Dolby Atmos) phá vỡ sự liên kết giữa nội dung và các vị trí loa cụ thể này. Thay vào đó, nội dung có thể được mô tả là tập hợp các đối tượng âm thanh riêng, mỗi đối tượng có thể có siêu dữ liệu thay đổi theo thời

gian mô tả vị trí cảm nhận mong muốn về các đối tượng âm thanh đã nêu trong không gian ba chiều. Tại thời điểm phát lại, nội dung được chuyển thành nguồn cấp qua loa bằng bộ kết xuất làm thích ứng với số lượng và vị trí loa trong hệ thống phát lại. Tuy nhiên, nhiều bộ kết xuất như vậy vẫn ràng buộc các vị trí của tập hợp loa là một trong số tập hợp bố trí được chỉ định (ví dụ 3.1.2, 5.1.2, 7.1.4, 9.1.6, v.v. với Dolby Atmos).

Vượt ra ngoài kết xuất ràng buộc như vậy, các phương pháp đã được phát triển cho phép âm thanh dựa trên đối tượng được kết xuất linh hoạt trên số lượng loa tùy ý thực sự đặt ở các vị trí tùy ý. Các phương pháp này yêu cầu bộ kết xuất phải có kiến thức về số lượng và vị trí vật lý của các loa trong không gian nghe. Để hệ thống như vậy trở nên thực tế đối với người tiêu dùng bình thường, mong muốn có phương pháp tự động để xác định vị trí của các loa. Một trong các phương pháp như vậy dựa vào việc sử dụng vô số micrô, có thể được bố trí cùng với loa. Khi phát tín hiệu âm thanh qua các loa và ghi bằng micrô, khoảng cách giữa mỗi loa và micrô được ước lượng. Từ những khoảng cách này, vị trí của cả loa và micrô sau đó được suy ra.

Đồng thời với sự ra đời của âm thanh không gian dựa trên đối tượng trong không gian tiêu dùng là sự áp dụng nhanh chóng của thiết bị gọi là “loa thông minh”, như dòng sản phẩm Echo của Amazon. Sự phổ biến rộng rãi của các thiết bị này có thể là do tính đơn giản và tiện lợi của chúng nhờ kết nối không dây và giao diện giọng nói tích hợp (ví dụ, Alexa của Amazon), nhưng khả năng âm thanh của các thiết bị này nhìn chung còn hạn chế, đặc biệt là đối với âm thanh không gian. Trong hầu hết các trường hợp, các thiết bị này bị ràng buộc phát lại đơn âm hoặc âm thanh nổi. Tuy nhiên, việc kết hợp các công nghệ kết xuất linh hoạt và định vị tự động nêu trên với nhiều loa thông minh hòa âm có thể tạo ra hệ thống có khả năng phát lại không gian rất phức tạp nhưng vẫn cực kỳ đơn giản cho người tiêu dùng thiết lập. Người tiêu dùng có thể đặt nhiều hoặc ít loa tùy thích, ở bất cứ đâu thấy thuận tiện mà không cần phải chạy dây loa do kết nối không dây, và micrô tích hợp có thể được dùng để tự động định vị các loa cho bộ kết xuất linh hoạt liên quan.

Các thuật toán kết xuất linh hoạt thông thường được thiết kế để đạt được ấn tượng không gian cảm nhận mong muốn cụ thể càng gần càng tốt. Trong hệ thống loa thông minh hòa âm, đôi khi, việc duy trì ấn tượng không gian này có thể không phải là mục tiêu quan trọng hoặc mong muốn nhất. Ví dụ, nếu ai đó đồng thời cố gắng nói với trợ lý

giọng nói tích hợp, thì có thể mong muốn thay đổi ngay lập tức kết xuất không gian theo cách giảm các mức phát lại tương đối trên các loa gần một số micrô nhất định để tăng tỷ lệ tín hiệu trên nhiễu và/hoặc tỷ lệ tín hiệu trên tiếng vọng (SER) của các tín hiệu micrô bao gồm lời nói được phát hiện. Một số phương án được mô tả ở đây có thể được thực hiện dưới dạng sửa đổi đối với các phương pháp kết xuất linh hoạt hiện có, để cho phép sửa đổi động việc kết xuất không gian như vậy, ví dụ, nhằm mục đích đạt được một hoặc nhiều mục đích bổ sung.

Các kỹ thuật kết xuất linh hoạt hiện có bao gồm Quét Biên độ trọng tâm (Center of Mass Amplitude Panning - CMAP) và Ảo hóa Linh hoạt (Flexible Virtualization - FV). Từ mức cao, cả hai kỹ thuật này kết xuất tập hợp một hoặc nhiều tín hiệu âm thanh, mỗi trong số chúng có một vị trí không gian nhận thức mong muốn liên quan, để phát lại trên tập hợp hai hoặc nhiều loa, trong đó việc kích hoạt tương đối các loa của tập hợp là một hàm của mô hình về vị trí không gian cảm nhận được của các tín hiệu âm thanh nêu trên được phát lại qua các loa và độ gần của vị trí không gian cảm nhận mong muốn của tín hiệu âm thanh đến các vị trí của các loa. Mô hình đảm bảo rằng người nghe có thể nghe thấy tín hiệu âm thanh ở gần vị trí không gian dự định của nó, và số hạng độ lân cận điều khiển các loa nào được dùng để đạt được ấn tượng không gian này. Cụ thể, số hạng độ lân cận giúp kích hoạt các loa ở gần vị trí không gian cảm nhận mong muốn của tín hiệu âm thanh. Đối với cả CMAP và FV, mỗi quan hệ hàm này được suy ra một cách thuận tiện từ hàm chi phí được viết dưới dạng tổng của hai số hạng, một số hạng đối với khía cạnh không gian và một số hạng đối với độ lân cận:

$$C(\mathbf{g}) = C_{\text{spatial}}(\mathbf{g}, \vec{o}, \{\vec{s}_i\}) + C_{\text{proximity}}(\mathbf{g}, \vec{o}, \{\vec{s}_i\}) \quad (1)$$

Ở đây, tập hợp $\{\vec{s}_i\}$ biểu thị các vị trí của tập hợp gồm M loa, $\vec{o}$ biểu thị vị trí không gian cảm nhận mong muốn của tín hiệu âm thanh, và $\mathbf{g}$ biểu thị vector M chiều để kích hoạt loa. Đối với CMAP, mỗi kích hoạt trong vector biểu diễn một độ lợi trên mỗi loa, còn đối với FV, mỗi kích hoạt biểu diễn bộ lọc (trong trường hợp thứ hai này, $\mathbf{g}$ có thể được coi là tương đương với vector của các giá trị phức ở một tần số cụ thể và $\mathbf{g}$ khác được tính toán trên nhiều tần số để tạo ra bộ lọc). Vector kích hoạt tối ưu được cho bằng cách tối thiểu hóa kích hoạt trên hàm chi phí:

$$\mathbf{g}_{\text{opt}} = \min_{\mathbf{g}} C(\mathbf{g}, \vec{o}, \{\vec{s}_i\}) \quad (2a)$$

Với một số định nghĩa nhất định của hàm chi phí, rất khó kiểm soát mức tuyệt đối của các kích hoạt tối ưu do việc tối thiểu hóa ở trên, mặc dù mức tương đối giữa các thành phần $\mathbf{g}_{opt}$ là thích hợp. Để giải quyết vấn đề này, việc chuẩn hóa tiếp theo $\mathbf{g}_{opt}$ có thể được thực hiện sao cho mức tuyệt đối của các kích hoạt được kiểm soát. Ví dụ, việc chuẩn hóa của vectơ để có độ dài đơn vị có thể được mong muốn, việc này phù hợp với các quy tắc quét công suất không đổi thường được dùng:

$$\bar{\mathbf{g}}_{opt} = \frac{\mathbf{g}_{opt}}{\|\mathbf{g}_{opt}\|} \quad (2b)$$

Hành vi chính xác của thuật toán kết xuất linh hoạt được quyết định bằng cấu trúc cụ thể của hai số hạng của hàm chi phí, $C_{spatial}$ và $C_{proximity}$. Đối với CMAP, $C_{spatial}$ được suy ra từ mô hình để đặt vị trí không gian cảm nhận được của tín hiệu âm thanh phát từ tập hợp các loa ở trọng tâm của các vị trí của các loa đó được lấy trọng số bằng độ lợi kích hoạt liên quan của chúng g_i (các phân tử của vectơ $\mathbf{g}$):

$$\vec{o} = \frac{\sum_{i=1}^M g_i \vec{s}_i}{\sum_{i=1}^M g_i} \quad (3)$$

Sau đó, phương trình 3 được xử lý thành chi phí không gian biểu diễn sai số bình phương giữa vị trí âm thanh mong muốn và vị trí được tạo ra bằng các loa kích hoạt:

$$C_{spatial}(\mathbf{g}, \vec{o}, \{\vec{s}_i\}) = \|(\sum_{i=1}^M g_i) \vec{o} - \sum_{i=1}^M g_i \vec{s}_i\|^2 = \|\sum_{i=1}^M g_i (\vec{o} - \vec{s}_i)\|^2 \quad (4)$$

Với FV, số hạng không gian của hàm chi phí được định nghĩa theo cách khác. Do đó mục tiêu là tạo ra đáp ứng âm thanh nổi $\mathbf{b}$ tương ứng với vị trí đối tượng âm thanh $\vec{o}$ ở tai trái và tai phải của người nghe. Về mặt khái niệm, $\mathbf{b}$ là vectơ 2×1 của các bộ lọc (một bộ lọc cho mỗi tai) nhưng được xử lý thuận tiện hơn dưới dạng vectơ 2×1 của các giá trị phức ở một tần số cụ thể. Tiếp theo dạng biểu diễn này ở một tần số cụ thể, đáp ứng âm thanh nổi mong muốn có thể được truy hồi từ tập hợp các HRTF được lập chỉ số theo vị trí đối tượng:

$$\mathbf{b} = HRTF\{\vec{o}\} \quad (5)$$

Đồng thời, đáp ứng âm thanh nổi 2×1 $\mathbf{e}$ tạo ra ở tai người nghe bằng các loa được lập mô hình dưới dạng ma trận truyền âm $2 \times M$ $\mathbf{H}$ nhân với vectơ $M \times 1$ $\mathbf{g}$ của các giá trị kích hoạt loa phức:

$$\mathbf{e} = \mathbf{H}\mathbf{g} \quad (6)$$

Ma trận truyền âm $\mathbf{H}$ được lập mô hình dựa trên tập hợp các vị trí loa $\{\vec{s}_i\}$ so với vị trí người nghe. Cuối cùng, thành phần không gian của hàm chi phí được định nghĩa là sai

số bình phương giữa đáp ứng âm thanh nổi mong muốn (phương trình 5) và đáp ứng do loa tạo ra (phương trình 6):

$$C_{spatial}(\mathbf{g}, \vec{o}, \{\vec{s}_i\}) = (\mathbf{b} - \mathbf{H}\mathbf{g})^*(\mathbf{b} - \mathbf{H}\mathbf{g}) \quad (7)$$

Có lợi là, số hạng không gian của hàm chi phí cho CMAP và FV được xác định trong các phương trình 4 và 7 đều có thể được sắp xếp lại thành một ma trận bậc hai là hàm kích hoạt loa $\mathbf{g}$:

$$C_{spatial}(\mathbf{g}, \vec{o}, \{\vec{s}_i\}) = \mathbf{g}^* \mathbf{A} \mathbf{g} + \mathbf{B} \mathbf{g} + \mathbf{C} \quad (8)$$

trong đó $\mathbf{A}$ là ma trận vuông $M \times M$, $\mathbf{B}$ là vector $1 \times M$ và $\mathbf{C}$ là đại lượng vô hướng. Ma trận $\mathbf{A}$ có hạng 2, và do đó khi $M > 2$, có tồn tại vô số lần kích hoạt loa $\mathbf{g}$ trong đó số hạng sai số không gian bằng 0. Đưa vào số hạng thứ hai của hàm chi phí, $C_{proximity}$, loại bỏ tính bất định này và thu được nghiệm riêng có các đặc tính có lợi về mặt cảm nhận so với các nghiệm có thể có khác. Đối với cả CMAP và FV, $C_{proximity}$ được cấu tạo sao cho việc kích hoạt các loa có vị trí $\vec{s}_i$ ở xa vị trí tín hiệu âm thanh mong muốn $\vec{o}$ bị phạt nhiều hơn việc kích hoạt các loa có vị trí gần với vị trí mong muốn. Cấu trúc này mang lại tập hợp kích hoạt loa tối ưu nhưng thừa thớt, trong đó chỉ những loa ở gần vị trí của tín hiệu âm thanh mong muốn được kích hoạt một cách đáng kể, và về mặt thực tế dẫn đến sự tái tạo không gian của tín hiệu âm thanh mạnh mẽ hơn về mặt cảm nhận đối với chuyển động của người nghe xung quanh tập hợp loa.

Để đạt được mục đích này, số hạng thứ hai của hàm chi phí, $C_{proximity}$ có thể được định nghĩa là tổng có trọng số khoảng cách của các giá trị tuyệt đối được lấy bình phương các kích hoạt loa. Số hạng này được biểu diễn rút gọn dưới dạng ma trận như sau:

$$C_{proximity}(\mathbf{g}, \vec{o}, \{\vec{s}_i\}) = \mathbf{g}^* \mathbf{D} \mathbf{g} \quad (9a)$$

trong đó $\mathbf{D}$ là ma trận đường chéo của các mức phạt khoảng cách giữa vị trí âm thanh mong muốn và mỗi loa:

$$\mathbf{D} = \begin{bmatrix} d_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & d_M \end{bmatrix}, \quad d_i = distance(\vec{o}, \vec{s}_i) \quad (9b)$$

Hàm phạt khoảng cách có thể có nhiều dạng, nhưng sau đây là dạng tham số hóa có ích

$$distance(\vec{o}, \vec{s}_i) = \alpha d_0^2 \left(\frac{\|\vec{o} - \vec{s}_i\|}{d_0} \right)^\beta \quad (9c)$$

trong đó $\|\vec{o} - \vec{s}_i\|$ là khoảng cách O-clit giữa vị trí âm thanh mong muốn và vị trí loa và α và β là các tham số điều chỉnh được. Tham số α biểu thị cường độ chung của giá trị phạt; d_0 tương ứng với phạm vi không gian của mức phạt khoảng cách (loa ở khoảng cách xấp xỉ d_0 hoặc xa hơn sẽ bị phạt), và β tính đến sự đột ngột của việc bắt đầu hình phạt ở khoảng cách d_0 .

Kết hợp hai số hạng của hàm chi phí được xác định trong phương trình 8 và phương trình 9a tạo ra hàm chi phí tổng thể

$$C(\mathbf{g}) = \mathbf{g}^* \mathbf{A} \mathbf{g} + \mathbf{B} \mathbf{g} + \mathbf{C} + \mathbf{g}^* \mathbf{D} \mathbf{g} = \mathbf{g}^* (\mathbf{A} + \mathbf{D}) \mathbf{g} + \mathbf{B} \mathbf{g} + \mathbf{C} \quad (10)$$

Đặt đạo hàm của hàm chi phí này với $\mathbf{g}$ bằng 0 và giải $\mathbf{g}$ sẽ thu được nghiệm kích hoạt loa tối ưu:

$$\mathbf{g}_{opt} = \frac{1}{2} (\mathbf{A} + \mathbf{D})^{-1} \mathbf{B} \quad (11)$$

Nói chung, nghiệm tối ưu trong phương trình 11 có thể tạo ra các kích hoạt loa có giá trị âm. Đối với cấu trúc CMAP của bộ kết xuất linh hoạt, các mức kích hoạt âm như vậy có thể không được mong muốn, và do đó phương trình (11) có thể được cực tiểu hóa tùy theo tất cả các mức kích hoạt vẫn còn dương.

Các hình vẽ trên Fig.15 và Fig.16 là sơ đồ minh họa ví dụ về tập hợp kích hoạt loa và các vị trí kết xuất đối tượng. Trong các ví dụ này, việc kích hoạt loa và vị trí kết xuất đối tượng tương ứng với các vị trí loa 4, 64, 165, -87 và -4 độ. Fig.15 thể hiện các kích hoạt loa 1505a, 1510a, 1515a, 1520a và 1525a, bao gồm nghiệm tối ưu cho phương trình 11 đối với các vị trí loa cụ thể này. Fig.16 vẽ biểu đồ các vị trí loa riêng là các điểm 1605, 1610, 1615, 1620 và 1625, lần lượt tương ứng với các kích hoạt loa 1505a, 1510a, 1515a, 1520a và 1525a. Fig.16 còn thể hiện các vị trí đối tượng lý tưởng (nói cách khác, các vị trí mà các đối tượng âm thanh sẽ được kết xuất tại đó) cho vô số góc đối tượng có thể là các điểm 1630a và vị trí kết xuất thực tế tương ứng cho các đối tượng đó là điểm 1635a, được kết nối với vị trí đối tượng lý tưởng bằng các đường chấm 1640a.

Một lớp phương án bao gồm các phương pháp kết xuất âm thanh để phát lại bằng ít nhất một (ví dụ, tất cả hoặc một số) trong số nhiều thiết bị âm thanh thông minh được phối hợp (được hòa âm). Ví dụ, tập hợp các thiết bị âm thanh thông minh có mặt (trong hệ thống) trong nhà của người dùng có thể được hòa âm để xử lý nhiều trường hợp sử dụng đồng thời, bao gồm kết xuất linh hoạt (theo một phương án) âm thanh để phát lại bằng tất cả hoặc một số (chẳng hạn, bằng (các) loa của tất cả hoặc một số) thiết bị âm

thanh thông minh. Nhiều tương tác với hệ thống được dự tính sẽ cần các thay đổi động cho việc kết xuất. Những thay đổi như vậy có thể, nhưng không nhất thiết, tập trung vào độ tin cậy theo không gian.

Một số phương án là các phương pháp kết xuất âm thanh để phát lại bằng ít nhất một (ví dụ, tất cả hoặc một số) thiết bị âm thanh thông minh của tập hợp các thiết bị âm thanh thông minh (hoặc để phát lại bằng ít nhất một (ví dụ, tất cả hoặc một số) loa của tập hợp loa khác). Việc kết xuất có thể bao gồm tối thiểu hóa hàm chi phí, trong đó hàm chi phí bao gồm ít nhất một số hạng kích hoạt loa động. Ví dụ về số hạng kích hoạt loa động như vậy bao gồm (nhưng không giới hạn ở):

- Độ gần của loa đến một hoặc nhiều người nghe;
- Độ gần của loa đến lực thu hút hoặc đẩy;
- Khả năng nghe được của loa đối với một số vị trí (ví dụ, vị trí người nghe, hoặc phòng em bé);
- Khả năng của loa (ví dụ, đáp ứng tần số và độ méo);
- Đồng bộ hóa các loa với các loa khác;
- Hiệu suất từ đánh thức; và
- Hiệu suất của bộ khử âm vọng.

(Các) số hạng kích hoạt loa động có thể cho phép ít nhất một trong số nhiều hành vi khác nhau, bao gồm việc làm lệch dạng biểu diễn không gian của âm thanh khỏi thiết bị âm thanh thông minh cụ thể để micrô của thiết bị này có thể nghe thấy người nói tốt hơn hoặc sao cho luồng âm thanh phụ có thể nghe tốt hơn từ (các) loa của thiết bị âm thanh thông minh.

Một số phương án thực hiện kết xuất để phát lại bằng (các) loa của nhiều thiết bị âm thanh thông minh được phối hợp (hòa âm). Các phương án khác thực hiện kết xuất để phát lại bằng (các) loa của tập hợp loa khác.

Việc ghép cặp các phương pháp kết xuất linh hoạt (được áp dụng theo một số phương án) với tập hợp loa thông minh không dây (hoặc các thiết bị âm thanh thông minh khác) có thể tạo ra hệ thống kết xuất âm thanh không gian cực kỳ có khả năng và dễ sử dụng. Khi dự tính các tương tác với hệ thống như vậy, hiển nhiên là các thay đổi động đối với kết xuất không gian có thể được mong muốn để tối ưu hóa cho các mục đích khác có thể phát sinh trong quá trình sử dụng hệ thống. Để đạt được mục tiêu này,

một lớp các phương án tăng cường các thuật toán kết xuất linh hoạt hiện có (trong đó kích hoạt loa là hàm của các số hạng không gian và độ gần mô tả trước đó), với một hoặc nhiều hàm có thể tạo cấu hình động bổ sung phụ thuộc vào một hoặc nhiều đặc tính của tín hiệu âm thanh đang được kết xuất, tập hợp loa và/hoặc các đầu vào bên ngoài khác. Theo một số phương án, hàm chi phí của kết xuất linh hoạt hiện có được xác định trong phương trình 1 được tăng cường với một hoặc nhiều phụ thuộc bổ sung theo

$$C(\mathbf{g}) = C_{spatial}(\mathbf{g}, \vec{o}, \{\vec{s}_i\}) + C_{proximity}(\mathbf{g}, \vec{o}, \{\vec{s}_i\}) + \sum_j C_j(\mathbf{g}, \{\{\hat{o}\}, \{\hat{s}_i\}, \{\hat{e}\}\}_j) \quad (12)$$

Trong phương trình 12, các số hạng $C_j(\mathbf{g}, \{\{\hat{o}\}, \{\hat{s}_i\}, \{\hat{e}\}\}_j)$ thể hiện các số hạng chi phí bổ sung, với $\{\hat{o}\}$ biểu diễn tập hợp gồm một hoặc nhiều đặc tính của tín hiệu âm thanh (ví dụ, chương trình âm thanh dựa trên đối tượng) đang được kết xuất, $\{\hat{s}_i\}$ biểu diễn tập hợp gồm một hoặc nhiều đặc tính của loa mà âm thanh đang được kết xuất và $\{\hat{e}\}$ biểu diễn một hoặc nhiều đầu vào bên ngoài bổ sung. Mỗi số hạng $C_j(\mathbf{g}, \{\{\hat{o}\}, \{\hat{s}_i\}, \{\hat{e}\}\}_j)$ trả về chi phí dưới dạng hàm kích hoạt $\mathbf{g}$ liên quan đến kết hợp của một hoặc nhiều đặc tính của tín hiệu âm thanh, loa, và/hoặc các đầu vào bên ngoài, được biểu diễn chung bằng tập hợp $\{\{\hat{o}\}, \{\hat{s}_i\}, \{\hat{e}\}\}_j$. Cần lưu ý rằng tập hợp $\{\{\hat{o}\}, \{\hat{s}_i\}, \{\hat{e}\}\}_j$ ở mức tối thiểu chỉ chứa một phần tử từ bất kỳ trong số $\{\hat{o}\}$, $\{\hat{s}_i\}$, hoặc $\{\hat{e}\}$.

Các ví dụ về $\{\hat{o}\}$ bao gồm, nhưng không giới hạn ở:

- Vị trí không gian cảm nhận mong muốn của tín hiệu âm thanh;
- Mức (có thể thay đổi theo thời gian) của tín hiệu âm thanh; và/hoặc
- Phổ (có thể thay đổi theo thời gian) của tín hiệu âm thanh.

Các ví dụ về $\{\hat{s}_i\}$ bao gồm, nhưng không giới hạn ở:

- Vị trí của loa trong không gian nghe;
- Đáp ứng tần số của loa;
- Giới hạn mức phát lại của loa;
- Các tham số của thuật toán xử lý động trong loa, như độ lợi của bộ giới hạn;
- Số đo hoặc ước lượng cuộc truyền âm thanh từ loa này đến những loa khác;
- Số đo hiệu suất của bộ khử tiếng vọng trên loa; và/hoặc

- Đồng bộ hóa tương đối của các loa với nhau.

Các ví dụ về $\{\varepsilon\}$ bao gồm, nhưng không giới hạn ở:

- Vị trí của một hoặc nhiều người nghe hoặc người nói trong không gian phát lại;
- Số đo hoặc ước lượng cuộc truyền âm thanh từ mỗi loa đến vị trí nghe;
- Số đo hoặc ước lượng cuộc truyền âm thanh từ người nói đến tập hợp loa;
- Vị trí của điểm mốc khác nào đó trong không gian phát lại; và/hoặc
- Số đo hoặc ước lượng cuộc truyền âm thanh từ mỗi loa đến điểm mốc khác nào đó trong không gian phát lại;

Với hàm chi phí mới được xác định trong phương trình 12, tập hợp các kích hoạt tối ưu có thể được tìm thấy nhờ việc tối thiểu hóa đối với $\mathbf{g}$ và chuẩn hóa sau có thể có như được xác định trước đó trong phương trình 2a và 2b.

Fig.17 là lưu đồ mô tả một ví dụ về phương pháp có thể được thực hiện bằng thiết bị hoặc hệ thống như thể hiện trên Fig.2A. Các khối của phương pháp 1700, tương tự các phương pháp khác được mô tả ở đây, không cần thiết được thực hiện theo thứ tự biểu thị. Hơn nữa, các phương pháp như vậy có thể bao gồm nhiều khối hoặc ít khối hơn so với phương pháp được thể hiện và/hoặc mô tả. Các khối của phương pháp 1700 có thể được thực hiện bằng một hoặc nhiều thiết bị, có thể là (hoặc có thể bao gồm) hệ thống điều khiển như hệ thống điều khiển 210 được thể hiện trên Fig.2A.

Trong phương án thực hiện này, khối 1705 bao gồm bước nhận dữ liệu âm thanh bằng hệ thống điều khiển và qua hệ thống giao diện. Trong ví dụ này, dữ liệu âm thanh bao gồm một hoặc nhiều tín hiệu âm thanh và dữ liệu không gian liên quan. Theo phương án thực hiện này, dữ liệu không gian biểu thị vị trí không gian cảm nhận được dự kiến tương ứng với tín hiệu âm thanh. Trong một số trường hợp, vị trí không gian cảm nhận được dự kiến có thể rõ ràng, ví dụ, như được biểu thị bằng siêu dữ liệu vị trí, như siêu dữ liệu vị trí Dolby Atmos. Trong các trường hợp khác, vị trí không gian cảm nhận được dự kiến có thể là ẩn, ví dụ, vị trí không gian cảm nhận được dự kiến có thể là vị trí giả định được liên kết với kênh theo Dolby 5.1, Dolby 7.1 hoặc định dạng âm thanh dựa trên kênh khác. Theo một số ví dụ, khối 1705 bao gồm modul kết xuất của hệ thống điều khiển nhận, dữ liệu âm thanh thông qua hệ thống giao diện.

Theo ví dụ này, khối 1710 bao gồm bước kết xuất, bằng hệ thống điều khiển, dữ liệu âm thanh để tái tạo qua tập hợp loa của môi trường, để tạo ra các tín hiệu âm thanh

được kết xuất. Trong ví dụ này, việc kết xuất mỗi trong số một hoặc nhiều tín hiệu âm thanh có trong dữ liệu âm thanh bao gồm bước xác định kích hoạt tương đối của tập hợp loa trong môi trường bằng cách tối ưu hóa hàm chi phí. Theo ví dụ này, chi phí là hàm của mô hình về vị trí không gian cảm nhận được của tín hiệu âm thanh khi được phát lại trên tập hợp loa trong môi trường. Trong ví dụ này, chi phí cũng là hàm của số đo độ gần của vị trí không gian cảm nhận được dự kiến của tín hiệu âm thanh với vị trí của mỗi loa trong tập hợp loa. Trong phương án thực hiện này, chi phí cũng là hàm của một hoặc nhiều hàm có thể tạo cấu hình động bổ sung. Trong ví dụ này, các hàm có thể tạo cấu hình động được dựa trên một hoặc nhiều yếu tố sau: độ gần của loa với một hoặc nhiều người nghe; độ gần của loa với vị trí lực hút, trong đó lực hút là yếu tố ưu tiên kích hoạt loa tương đối cao hơn ở khoảng cách gần với vị trí lực hút hơn; độ gần của loa với vị trí lực đẩy, trong đó lực đẩy là yếu tố ưu tiên kích hoạt loa thấp hơn tương đối ở khoảng cách gần vị trí lực đẩy hơn; khả năng của mỗi loa so với các loa khác trong môi trường; sự đồng bộ hóa của các loa với các loa khác; hiệu suất từ đánh thức; hoặc hiệu suất của bộ khử tiếng vọng.

Trong ví dụ này, khối 1715 bao gồm bước cung cấp, qua hệ thống giao diện, các tín hiệu âm thanh được kết xuất cho ít nhất một số loa trong tập hợp loa của môi trường.

Theo một số ví dụ, mô hình về vị trí không gian cảm nhận được có thể tạo ra đáp ứng âm thanh nổi tương ứng với vị trí đối tượng âm thanh ở tai trái và tai phải của người nghe. Ngoài ra, hoặc cách khác, mô hình về vị trí không gian cảm nhận được có thể đặt vị trí không gian cảm nhận được của tín hiệu âm thanh phát từ tập hợp loa tại trọng tâm của tập hợp vị trí của các loa có trọng số bằng độ lợi kích hoạt liên quan của loa.

Theo một số ví dụ, một hoặc nhiều hàm có thể tạo cấu hình động bổ sung có thể dựa, ít nhất một phần, vào mức của một hoặc nhiều tín hiệu âm thanh. Trong một số trường hợp, một hoặc nhiều hàm tạo cấu hình động bổ sung có thể dựa, ít nhất một phần, vào phổ của một hoặc nhiều tín hiệu âm thanh.

Một số ví dụ về phương pháp 1700 bao gồm bước nhận thông tin bố trí loa. Theo một số ví dụ, một hoặc nhiều hàm có thể tạo cấu hình động bổ sung có thể dựa, ít nhất một phần, vào vị trí của mỗi loa trong môi trường.

Một số ví dụ về phương pháp 1700 bao gồm bước nhận thông tin đặc điểm kỹ thuật của loa. Theo một số ví dụ, một hoặc nhiều hàm có thể tạo cấu hình động bổ sung

có thể dựa, ít nhất một phần, vào khả năng của mỗi loa, có thể bao gồm một hoặc nhiều trong số đáp ứng tần số, giới hạn mức phát lại hoặc các tham số của một hoặc nhiều thuật toán xử lý động lực học của loa.

Theo một số ví dụ, một hoặc nhiều hàm có thể tạo cấu hình động bổ sung có thể dựa, ít nhất một phần, vào số đo hoặc ước lượng cuộc truyền âm thanh từ loa này đến các loa khác. Cách khác, hoặc ngoài ra, một hoặc nhiều hàm có thể tạo cấu hình động bổ sung có thể dựa, ít nhất một phần, vào vị trí người nghe hoặc loa của một hoặc nhiều người trong môi trường. Cách khác, hoặc ngoài ra, một hoặc nhiều hàm có thể tạo cấu hình động bổ sung có thể dựa, ít nhất một phần, vào số đo hoặc ước lượng cuộc truyền âm thanh từ mỗi loa đến người nghe hoặc vị trí của loa. Ví dụ, việc ước lượng về cuộc truyền âm có thể dựa ít nhất một phần vào tường, đồ nội thất hoặc các vật thể khác có thể nằm giữa mỗi loa và người nghe hoặc vị trí của loa.

Ngoài ra, hoặc cách khác, một hoặc nhiều hàm có thể tạo cấu hình động bổ sung có thể dựa, ít nhất một phần, vào vị trí đối tượng của một hoặc nhiều đối tượng hoặc điểm mốc không phải loa trong môi trường. Theo một số phương án thực hiện như vậy, một hoặc nhiều hàm có thể tạo cấu hình động bổ sung có thể dựa, ít nhất một phần, vào số đo hoặc ước lượng cuộc truyền âm thanh từ mỗi loa đến vị trí đối tượng hoặc vị trí mốc.

Nhiều hành vi mới và hữu ích có thể đạt được bằng cách sử dụng một hoặc nhiều số hạng chi phí bổ sung được xác định phù hợp để thực hiện kết xuất linh hoạt. Tất cả các hành vi làm ví dụ được liệt kê dưới đây được áp dụng để xử phạt một số loa nhất định trong một số điều kiện nhất định được coi là không mong muốn. Kết quả cuối cùng là các loa này được kích hoạt ít hơn trong quy trình kết xuất không gian của tập hợp tín hiệu âm thanh. Trong nhiều trường hợp như vậy, sáng chế có thể dự tính đến việc giảm âm lượng các loa không mong muốn độc lập với việc thay đổi bất kỳ đối với kết xuất không gian, nhưng chiến lược như vậy có thể làm suy giảm đáng kể sự cân bằng tổng thể của nội dung âm thanh. Ví dụ, một số thành phần của hỗn hợp có thể trở nên hoàn toàn không nghe được. Mặt khác, với các phương án được mô tả, việc tích hợp các hình phạt này vào quy trình tối ưu hóa lõi của kết xuất cho phép kết xuất thích ứng và thực hiện kết xuất không gian tốt nhất có thể với các loa ít bị phạt còn lại. Đây là một giải pháp khéo léo, dễ thích nghi, và hiệu quả hơn nhiều.

Ví dụ về các trường hợp sử dụng bao gồm, nhưng không giới hạn:

- Tạo ra dạng biểu diễn không gian cân bằng hơn quanh khu vực nghe
 - Thấy rằng âm thanh không gian được trình bày tốt nhất trên các loa cơ bản có cùng khoảng cách với khu vực nghe dự kiến. Chi phí có thể được xây dựng sao cho các loa gần hoặc xa hơn đáng kể so với khoảng cách trung bình của loa đến khu vực nghe bị phạt, do đó giảm được quy trình kích hoạt của chúng;
- Di chuyển âm thanh ra xa hoặc về phía người nghe hoặc người nói
 - Nếu người dùng hệ thống cố gắng nói với trợ lý giọng nói thông minh của hoặc được kết hợp với hệ thống, thì việc tạo ra chi phí để phạt các loa gần người nói hơn có thể có lợi. Bằng cách này, các loa này được kích hoạt ít hơn, cho phép các micrô liên quan của chúng nghe thấy người nói tốt hơn;
 - Để tạo ra trải nghiệm quen thuộc hơn cho người nghe để tối thiểu hóa các mức phát lại cho những người khác trong không gian nghe, các loa ở xa vị trí của người nghe có thể bị phạt nặng để chỉ những loa gần người nghe nhất mới được kích hoạt;
- Di chuyển âm thanh ra xa hoặc về phía một điểm mốc, vùng hoặc khu vực
 - Một số vị trí nhất định gần không gian nghe có thể được coi là nhạy cảm, như phòng của em bé, giường của em bé, phòng làm việc, khu vực đọc sách, khu vực học tập, v.v. Trong trường hợp này, chi phí có thể được xây dựng để phạt việc sử dụng loa gần vị trí, vùng hoặc khu vực này;
 - Ngoài ra, đối với trường hợp như trên (hoặc các trường hợp tương tự), hệ thống loa có thể đã tạo ra các số đo cuộc truyền âm thanh từ mỗi loa vào phòng của em bé, đặc biệt nếu một trong các loa (có micrô gắn kèm hoặc kết hợp) nằm trong chính phòng của bé. Trong trường hợp này, thay vì sử dụng khoảng cách vật lý của loa với phòng của em bé, chi phí có thể được xây dựng để phạt việc sử dụng các loa có cuộc truyền âm thanh đo được của chúng vào phòng là cao; và/hoặc
- Sử dụng tối ưu các khả năng của loa
 - Khả năng của các loa khác nhau có thể thay đổi đáng kể. Ví dụ, một loa thông minh phổ biến chỉ chứa một trình điều khiển toàn dải 1,6” với khả

năng tần số thấp giới hạn. Mặt khác, loa thông minh khác chứa loa trầm 3" có khả năng hơn nhiều. Những khả năng này thường được phản ánh trong đáp ứng tần số của loa, và như vậy, tập hợp các đáp ứng liên quan đến loa có thể được dùng trong số hạng chi phí. Ở tần số cụ thể, các loa có khả năng kém hơn so với các loa khác, được đo bằng đáp ứng tần số của chúng, có thể bị phạt và do đó được kích hoạt ở mức độ thấp hơn. Theo một số phương án thực hiện, các giá trị đáp ứng tần số như vậy có thể được lưu trữ bằng loa thông minh và sau đó được báo cáo cho đơn vị tính toán chịu trách nhiệm tối ưu hóa quy trình kết xuất linh hoạt;

- Nhiều loa chứa nhiều hơn một trình điều khiển, mỗi trình điều khiển chịu trách nhiệm để phát một dải tần số khác nhau. Ví dụ, một loa thông minh phổ biến có thiết kế hai đường tiếng chứa loa trầm cho tần số thấp hơn và loa cao tần cho tần số cao hơn. Thông thường, loa như vậy chứa mạch phân tần để chia tín hiệu âm thanh phát lại toàn dải thành các dải tần số thích hợp và gửi cho các trình điều khiển tương ứng. Theo cách khác, loa như vậy có thể cung cấp quyền truy cập phát lại bộ kết xuất linh hoạt cho từng trình điều khiển riêng cũng như thông tin về khả năng của từng trình điều khiển riêng, như đáp ứng tần số. Bằng cách áp dụng số hạng chi phí như được mô tả ở trên, theo một số ví dụ, bộ kết xuất linh hoạt có thể tự động xây dựng sự phân tần giữa hai trình điều khiển dựa trên khả năng tương đối của chúng ở các tần số khác nhau;
- Ví dụ được mô tả ở trên sử dụng đáp ứng tần số tập trung vào các khả năng vốn có của loa nhưng có thể không phản ánh chính xác khả năng của loa như khi được đặt trong môi trường nghe. Trong một số trường hợp nhất định, các đáp ứng tần số của loa được đo ở vị trí nghe dự kiến có thể khả dụng thông qua một số thủ tục hiệu chỉnh. Các phép đo như vậy có thể được dùng thay vì các đáp ứng được tính toán trước để tối ưu hóa tốt hơn việc sử dụng các loa. Ví dụ, loa nhất định có thể rất có khả năng ở một tần số cụ thể, nhưng do vị trí của nó (ví dụ như sau bức tường hoặc một phần đồ nội thất) có thể tạo ra đáp ứng rất hạn chế tại vị trí nghe dự

định. Phép đo ghi nhận đáp ứng này và được đưa vào số hạng chi phí thích hợp có thể ngăn chặn sự kích hoạt đáng kể của loa như vậy;

- Đáp ứng tần số chỉ là một khía cạnh của khả năng phát lại của loa. Nhiều loa nhỏ hơn bắt đầu bị méo và sau đó đạt đến giới hạn hành trình của chúng khi mức phát lại tăng lên, đặc biệt là đối với các tần số thấp hơn. Để giảm sự méo như vậy, nhiều loa thực hiện xử lý động lực học để ràng buộc mức phát lại thấp hơn một số ngưỡng giới hạn mà có thể thay đổi theo tần số. Trong các trường hợp loa ở gần hoặc tại các ngưỡng này, còn những loa khác tham gia kết xuất linh hoạt thì không, cần giảm mức tín hiệu trong loa giới hạn và chuyển hướng năng lượng này sang các loa khác ít bị phạt hơn. Hành vi như vậy có thể tự động đạt được theo một số phương án bằng cách tạo cấu hình chính xác số hạng chi phí liên quan. Số hạng chi phí như vậy có thể bao gồm một hoặc nhiều trong số:
 - Giám sát âm lượng phát lại toàn cục liên quan đến ngưỡng giới hạn của các loa. Ví dụ, loa có mức âm lượng gần với ngưỡng giới hạn hơn có thể bị phạt nhiều hơn;
 - Giám sát các mức tín hiệu động, có thể thay đổi theo tần số, liên quan đến ngưỡng giới hạn của loa, cũng có thể thay đổi theo tần số. Ví dụ, loa có mức tín hiệu được giám sát gần với ngưỡng giới hạn của nó hơn có thể bị phạt nhiều hơn;
 - Giám sát các tham số của quy trình xử lý động lực học của loa một cách trực tiếp, như giới hạn độ lợi. Theo một số ví dụ như vậy, loa có các tham số biểu thị giới hạn cao hơn có thể bị phạt nhiều hơn; và/hoặc
 - Theo dõi điện áp, dòng điện và công suất tức thời thực tế được bộ khuếch đại phân phối cho loa để xác định xem loa có đang hoạt động trong dải tuyến tính hay không. Ví dụ, loa hoạt động tuyến tính ít hơn có thể bị phạt nhiều hơn;
- Các loa thông minh có các micrô tích hợp và trợ lý giọng nói tương tác thường sử dụng dạng kỹ thuật khử tiếng vọng nào đó để giảm mức tín hiệu âm thanh phát ra từ loa khi thu được bằng micrô thu. Mức giảm này

càng lớn, loa càng có cơ hội nghe và hiểu người nói trong không gian tốt hơn. Nếu lượng dư của bộ khử tiếng vọng luôn ở mức cao, thì đây có thể là dấu hiệu thể hiện loa đang được điều khiển vào vùng phi tuyến tính trong đó việc dự đoán đường đi của tiếng vọng trở nên khó khăn. Trong trường hợp như vậy, có thể hợp lý khi chuyển hướng năng lượng tín hiệu ra khỏi loa và do đó, số hạng chi phí có tính đến hiệu suất của bộ khử tiếng vọng có thể có lợi. Số hạng chi phí như vậy có thể gán chi phí cao cho loa có bộ khử tiếng vọng gắn kèm của nó hoạt động kém;

- Để đạt được sự tạo ảnh có thể dự đoán khi kết xuất âm thanh không gian qua nhiều loa, thông thường, việc phát lại trên tập hợp loa phải được đồng bộ hóa hợp lý theo thời gian. Đối với loa có dây, việc này được xác định, nhưng với vô số loa không dây, việc đồng bộ hóa có thể là một thách thức và kết quả cuối cùng có thể thay đổi. Trong trường hợp như vậy, mỗi loa có thể báo cáo mức độ đồng bộ hóa tương đối của nó với đích, và mức này sau đó có thể được đưa vào số hạng chi phí đồng bộ hóa. Theo một số ví dụ như vậy, loa có mức độ đồng bộ hóa thấp hơn có thể bị phạt nhiều hơn và do đó bị loại khỏi quy trình kết xuất. Ngoài ra, có thể không cần đồng bộ hóa chặt chẽ đối với một số loại tín hiệu âm thanh, ví dụ như các thành phần của hỗn hợp âm thanh nhằm mục đích khuếch tán hoặc không định hướng. Theo một số phương án thực hiện, các thành phần có thể được gán thể như vậy với siêu dữ liệu và số hạng chi phí đồng bộ hóa có thể được thay đổi để giảm mức phạt.

Tiếp theo, sáng chế mô tả các ví dụ bổ sung của các phương án. Tương tự như chi phí độ gần được xác định trong phương trình 9a và 9b, cũng có thể thuận tiện để biểu thị mỗi trong số các số hạng hàm chi phí mới $C_j(\mathbf{g}, \{\{\hat{o}\}, \{\hat{s}_i\}, \{\hat{e}\}\}_j)$ dưới dạng tổng có trọng số của các giá trị tuyệt đối được lấy bình phương của các kích hoạt loa, ví dụ, như sau:

$$C_j(\mathbf{g}, \{\{\hat{o}\}, \{\hat{s}_i\}, \{\hat{e}\}\}_j) = \mathbf{g}^* \mathbf{W}_j(\{\{\hat{o}\}, \{\hat{s}_i\}, \{\hat{e}\}\}_j) \mathbf{g}, \quad (13a)$$

trong đó $\mathbf{W}_j$ là ma trận đường chéo của các trọng số $w_{ij} = w_{ij}(\{\{\hat{o}\}, \{\hat{s}_i\}, \{\hat{e}\}\}_j)$ mô tả chi phí liên quan đến việc kích hoạt loa i cho số hạng j :

$$\mathbf{W}_j = \begin{bmatrix} w_{1j} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_{Mj} \end{bmatrix} \quad (13b)$$

Kết hợp phương trình 13a và 13b với phiên bản ma trận bậc hai của hàm chi phí CMAP và FV được xác định trong phương trình 10 tạo ra phương án thực hiện có lợi có thể có của hàm chi phí mở rộng chung (theo một số phương án) được xác định trong phương trình 12:

$$C(\mathbf{g}) = \mathbf{g}^* \mathbf{A} \mathbf{g} + \mathbf{B} \mathbf{g} + \mathbf{C} + \mathbf{g}^* \mathbf{D} \mathbf{g} + \sum_j \mathbf{g}^* \mathbf{W}_j \mathbf{g} = \mathbf{g}^* (\mathbf{A} + \mathbf{D} + \sum_j \mathbf{W}_j) \mathbf{g} + \mathbf{B} \mathbf{g} + \mathbf{C} \quad (14)$$

Với định nghĩa này về các số hạng hàm chi phí mới, hàm chi phí tổng thể vẫn là ma trận bậc hai, và tập hợp kích hoạt tối ưu $\mathbf{g}_{opt}$ có thể được tìm thấy nhờ phép lấy vi phân của phương trình 14 để tạo ra

$$\mathbf{g}_{opt} = \frac{1}{2} (\mathbf{A} + \mathbf{D} + \sum_j \mathbf{W}_j)^{-1} \mathbf{B} \quad (15)$$

Có ích nếu coi mỗi số hạng trong số các số hạng trọng số w_{ij} là các hàm của giá trị phạt liên tục đã cho $p_{ij} = p_{ij}(\{\{\hat{o}\}, \{\hat{s}_i\}, \{\hat{e}\}\}_j)$ đối với mỗi loa trong số các loa. Theo một phương án ví dụ, giá trị phạt này là khoảng cách từ đối tượng (cần được kết xuất) đến loa được xem xét. Theo một phương án làm ví dụ khác, giá trị phạt này thể hiện việc không có khả năng tái tạo một số tần số của loa đã cho. Dựa trên giá trị phạt này, các số hạng trọng số w_{ij} có thể được tham số hóa dưới dạng:

$$w_{ij} = \alpha_j f_j \left(\frac{p_{ij}}{\tau_j} \right) \quad (16)$$

trong đó α_j biểu diễn hệ số trước (có tính đến cường độ chung của số hạng trọng số), trong đó τ_j biểu diễn ngưỡng phạt (xấp xỉ hoặc cao hơn ngưỡng này, số hạng trọng số trở nên có nghĩa), và trong đó $f_j(x)$ biểu diễn hàm tăng đơn điệu. Ví dụ, với $f_j(x) = x^{\beta_j}$ số hạng trọng số có dạng:

$$w_{ij} = \alpha_j \left(\frac{p_{ij}}{\tau_j} \right)^{\beta_j} \quad (17)$$

trong đó $\alpha_j, \beta_j, \tau_j$ là các tham số có thể điều chỉnh mà biểu thị lần lượt cường độ chung của giá trị phạt, tính đột ngột của việc bắt đầu phạt và phạm vi của mức phạt. Cần thận trọng khi thiết lập các giá trị có thể điều chỉnh này sao cho việc ảnh hưởng tương đối của số hạng chi phí C_j đối với các số hạng chi phí bổ sung khác bất kỳ cũng như $C_{spatial}$ và $C_{proximity}$ là thích hợp để đạt được kết quả mong muốn. Ví dụ, theo quy tắc chung, nếu một người muốn một hình phạt cụ thể để có ảnh hưởng rõ ràng đến những loa khác thì việc thiết lập cường độ của nó α_j lớn hơn gấp mười lần so với cường độ phạt lớn nhất tiếp theo có thể thích hợp.

Trong trường hợp tắt cả các loa đều bị phạt, thường dễ dàng trừ đi mức phạt tối thiểu cho tất cả các số hạng trọng số trong quá trình hậu xử lý để ít nhất một trong các loa không bị phạt:

$$w_{ij} \rightarrow w'_{ij} = w_{ij} - \min_i(w_{ij}) \quad (18)$$

Như đã nêu ở trên, có thể có nhiều trường hợp sử dụng có thể được thực hiện bằng cách sử dụng các số hạng hàm chi phí mới được mô tả ở đây (và các số hạng hàm chi phí mới tương tự được dùng theo các phương án khác). Tiếp theo, sáng chế mô tả chi tiết cụ thể hơn với ba ví dụ: di chuyển âm thanh về phía người nghe hoặc người nói, di chuyển âm thanh ra xa người nghe hoặc người nói, và di chuyển âm thanh ra xa điểm mốc.

Trong ví dụ thứ nhất, thuật ngữ được gọi ở đây là “lực hút” được dùng để hút âm thanh đến một vị trí, theo một số ví dụ có thể là vị trí của người nghe hoặc người nói, vị trí mốc, vị trí đồ nội thất, v.v. Vị trí này có thể được gọi ở đây là “vị trí lực hút” hoặc “vị trí hút”. Như được dùng ở đây, “lực hút” là yếu tố ưu tiên việc kích hoạt loa tương đối cao hơn ở gần vị trí lực hút hơn. Theo ví dụ này, trọng số w_{ij} có dạng trong phương trình 17 với giá trị phạt liên tục p_{ij} được xác định bằng khoảng cách của loa thứ i từ một vị trí thu hút cố định $\vec{l}_j$ và giá trị ngưỡng τ_j được xác định bằng giá trị lớn nhất của các khoảng cách này trên tất cả các loa:

$$p_{ij} = \|\vec{l}_j - \vec{s}_i\|, \text{ và} \quad (19a)$$

$$\tau_j = \max_i \|\vec{l}_j - \vec{s}_i\| \quad (19b)$$

Để minh họa trường hợp sử dụng “kéo” âm thanh về phía người nghe hoặc người nói, sáng chế thiết lập cụ thể $\alpha_j = 20$, $\beta_j = 3$, và $\vec{l}_j$ thành vector tương ứng với vị trí của

người nghe/người nói bằng 180 độ (phía dưới, tâm của đồ thị). Các giá trị α_j , β_j , và $\vec{l}_j$ này chỉ là ví dụ. Theo một số phương án thực hiện, α_j có thể nằm trong khoảng từ 1 đến 100 và β_j có thể nằm trong khoảng từ 1 đến 25. Fig.18 là đồ thị kích hoạt loa theo một phương án làm ví dụ. Trong ví dụ này, Fig.18 thể hiện các kích hoạt loa 1505b, 1510b, 1515b, 1520b và 1525b, bao gồm nghiệm tối ưu cho hàm chi phí cho các vị trí loa giống nhau từ Fig.15 và Fig.16, với việc bổ sung lực hút được biểu diễn bằng w_{ij} . Fig.19 là biểu đồ của các vị trí kết xuất đối tượng theo một phương án làm ví dụ. Trong ví dụ này, Fig.19 thể hiện vị trí đối tượng lý tưởng tương ứng 1630b cho vô số góc đối tượng có thể có và vị trí kết xuất thực tế tương ứng 1635b cho các đối tượng đó, được kết nối với các vị trí đối tượng lý tưởng 1630b bằng các đường chấm 1640b. Hướng xiên của các vị trí kết xuất thực tế 1635b về phía vị trí cố định $\vec{l}_j$ minh họa sự ảnh hưởng của các trọng số hút đối với nghiệm tối ưu cho hàm chi phí.

Trong ví dụ thứ hai và thứ ba, “lực đẩy” được dùng để “đẩy” âm thanh ra khỏi vị trí, có thể là vị trí của người (ví dụ, vị trí người nghe, vị trí người nói, v.v.) hoặc vị trí khác, như vị trí mốc, vị trí đồ nội thất, v.v. Theo một số ví dụ, lực đẩy có thể được dùng để đẩy âm thanh ra khỏi khu vực hoặc vùng của môi trường nghe, như khu vực văn phòng, khu vực đọc sách, khu vực giường ngủ hoặc phòng ngủ (ví dụ, giường hoặc phòng ngủ của em bé), v.v. Theo một số ví dụ như vậy, vị trí cụ thể có thể được dùng làm đại diện của một vùng hoặc một khu vực. Ví dụ, vị trí đại diện cho giường của em bé có thể là vị trí được ước lượng của đầu em bé, vị trí nguồn âm thanh được ước lượng tương ứng với em bé, v.v. Vị trí này có thể được gọi ở đây là “vị trí lực đẩy” hoặc “vị trí đẩy.” Như được dùng ở đây, “lực đẩy” là yếu tố ưu tiên việc kích hoạt loa tương đối thấp hơn ở gần vị trí lực đẩy hơn. Theo ví dụ này, sáng chế định nghĩa p_{ij} và τ_j đối với vị trí đẩy cố định $\vec{l}_j$ tương tự như lực hút trong phương trình 19:

$$p_{ij} = \max_i \|\vec{l}_j - \vec{s}_i\| - \|\vec{l}_j - \vec{s}_i\|, \text{ và} \quad (19c)$$

$$\tau_j = \max_i \|\vec{l}_j - \vec{s}_i\| \quad (19d)$$

Để minh họa trường hợp sử dụng đẩy âm thanh ra xa người nghe hoặc người nói, theo một ví dụ, sáng chế có thể thiết lập cụ thể $\alpha_j = 5$, $\beta_j = 2$, và $\vec{l}_j$ thành một vector tương ứng với vị trí người nghe/người nói bằng 180 độ (ở dưới cùng, tâm của đồ thị). Các giá

trị α_j , β_j , và $\vec{l}_j$ này chỉ là các ví dụ. Như lưu ý ở trên, theo một số ví dụ α_j có thể nằm trong khoảng từ 1 đến 100 và β_j có thể nằm trong khoảng từ 1 đến 25. Fig.20 là biểu đồ kích hoạt loa theo một phương án làm ví dụ. Theo ví dụ này, Fig.20 thể hiện các kích hoạt loa 1505c, 1510c, 1515c, 1520c và 1525c, bao gồm nghiệm tối ưu cho hàm chi phí cho các vị trí loa giống như các hình vẽ trước, với việc bổ sung lực đẩy được biểu diễn bằng w_{ij} . Fig.21 là biểu đồ của các vị trí kết xuất đối tượng theo một phương án làm ví dụ. Trong ví dụ này, Fig.21 thể hiện các vị trí đối tượng lý tưởng 1630c cho vô số góc đối tượng có thể có và các vị trí kết xuất thực tế tương ứng 1635c cho các đối tượng đó, được kết nối với các vị trí đối tượng lý tưởng 1630c bằng các đường chấm 1640c. Hướng xiên của các vị trí kết xuất thực tế 1635c cách vị trí cố định $\vec{l}_j$ minh họa tác động của các trọng số của bộ đẩy đối với nghiệm tối ưu cho hàm chi phí.

Trường hợp sử dụng làm ví dụ thứ ba là “đẩy” âm thanh ra xa điểm mốc nhạy cảm về âm thanh, như cửa vào phòng ngủ của em bé. Tương tự với ví dụ ngay trên, sáng chế đặt $\vec{l}_j$ thành một vectơ tương ứng với vị trí của 180 độ (dưới cùng, tâm của đồ thị). Để đạt được lực đẩy mạnh hơn và làm xiên hoàn toàn trường âm vào phần trước của không gian nghe chính, sáng chế đặt $\alpha_j = 20$, $\beta_j = 5$. Fig.22 là biểu đồ kích hoạt loa theo một phương án làm ví dụ. Lần nữa, trong ví dụ này, Fig.22 thể hiện các kích hoạt loa 1505d, 1510d, 1515d, 1520d và 1525d, bao gồm nghiệm tối ưu cho cùng một tập hợp vị trí loa với việc bổ sung lực đẩy mạnh hơn. Fig.23 là biểu đồ của các vị trí kết xuất đối tượng theo một phương án làm ví dụ. Và lần nữa, trong ví dụ này, Fig.23 thể hiện vị trí đối tượng lý tưởng 1630d cho vô số góc đối tượng có thể có và vị trí kết xuất thực tế tương ứng 1635d cho các đối tượng đó, được kết nối với vị trí đối tượng lý tưởng 1630d bằng các đường chấm 1640d. Hướng xiên của các vị trí kết xuất thực tế 1635d minh họa tác động của các trọng số bộ đẩy mạnh hơn đối với nghiệm tối ưu cho hàm chi phí.

Theo một ví dụ khác về phương pháp 250 trên Fig.2B, trường hợp sử dụng để đáp lại việc lựa chọn của hai hoặc nhiều thiết bị âm thanh trong môi trường âm thanh (khối 265) và áp dụng lực “đẩy” cho âm thanh (khối 275). Theo ví dụ trước đó, bước lựa chọn hai hoặc nhiều thiết bị âm thanh theo một số phương án có thể có dạng giá trị f_n , các tham số không có đơn vị để kiểm soát mức độ các thay đổi trong xử lý âm thanh xảy ra. Nhiều sự kết hợp là có thể. Theo một ví dụ đơn giản, các trọng số tương ứng với

lực đẩy có thể được lựa chọn trực tiếp là $w_{ij} = \alpha_j f_i^{\beta_j}$, phạt các thiết bị được chọn theo khía cạnh “quyết định”.

Ngoài các ví dụ trước đây về việc xác định trọng số, theo một số phương án thực hiện, trọng số có thể được xác định như sau:

$$w_{ij} = \alpha_j \left(f_i + \left(\frac{p_{ij}}{\tau_j} \right) \right)^{\beta_j}$$

Trong phương trình trên, α_j , β_j , τ_j biểu diễn các tham số có thể điều chỉnh được lần lượt biểu thị cường độ toàn cục của mức phạt, tính đột ngột khi bắt đầu phạt và mức độ phạt, như được mô tả ở trên liên quan đến phương trình 17. Do đó, phương trình trên có thể được hiểu là sự kết hợp của nhiều số hạng phạt, phát sinh từ nhiều trường hợp sử dụng đồng thời. Ví dụ, âm thanh có thể bị “đẩy ra” xa mốc nhạy cảm bằng cách sử dụng các số hạng p_{ij} và τ_j như được mô tả trong ví dụ trên đây, trong khi vẫn bị “đẩy ra” khỏi vị trí micrô khi có mong muốn SER được cải thiện bằng cách sử dụng số hạng f_i được xác định bằng khía cạnh quyết định.

Các ví dụ trước đây cũng đã đưa vào s_n , được biểu diễn trực tiếp bằng cải thiện tỷ lệ lời nói trên tiếng vọng theo decibels. Một số phương án có thể bao gồm bước lựa chọn các giá trị α và β (cường độ phạt, và độ đột ngột của thời điểm bắt đầu hình phạt tương ứng) dựa một phần trên giá trị s_n tính bằng dB, và các công thức được biểu thị trước đó cho w_{ij} có thể sử dụng α_{ij} và β_{ij} lần lượt thay cho α_j và β_j . Ví dụ, giá trị $s_i = -20$ dB sẽ tương ứng với chi phí cao để kích hoạt loa thứ i . Theo một số phương án như vậy, α_{ij} có thể được thiết lập thành số cao hơn nhiều lần so với các giá trị điển hình của các số hạng khác trong hàm chi phí $C_{spatial}$ và $C_{proximity}$. Ví dụ, giá trị alpha mới có thể được xác định bằng $\alpha'_{ij} = \alpha_j * 10^{-\frac{s_i}{20}}$, trong đó giá trị $s_i = -20$ dB, sẽ dẫn đến việc α_{ij} có giá trị lớn hơn 10 lần so với giá trị thường có trong hàm chi phí. Trong một số trường hợp, việc thay đổi β_{ij} sẽ được thiết lập trong khoảng $0,5 < \beta_{ij} < 1,0$ có thể là một thay đổi thích hợp dựa trên các giá trị âm lớn s_i , “đẩy” âm thanh ra xa vùng thậm chí còn lớn hơn xung quanh loa thứ i . Ví dụ, các giá trị s_i có thể được ánh xạ thành β_{ij} theo:

$$\beta_{ij} = \max \left(0,5, \min \left(1,0, 1,5 + \frac{s_i}{30,0} \right) \right)$$

Trong ví dụ này, đối với $s_i = -20,0$ dB, β_{ij} sẽ bằng 0,8333.

Các khía cạnh của các phương án làm ví dụ bao gồm các phương án làm ví dụ được đánh số sau đây (“enumerated example embodiment - EEE”):

EEE1. Phương pháp (hoặc hệ thống) để cải thiện tỷ lệ tín hiệu trên tiếng vọng để phát hiện các lệnh giọng nói từ người dùng, trong đó

- a. Có nhiều thiết bị được dùng để tạo ra dữ liệu chương trình âm thanh đầu ra
- b. Có tập hợp khoảng cách đã biết hoặc mối liên quan có thứ tự đối với các thiết bị từ người nghe
- c. Hệ thống giảm có lựa chọn âm lượng của thiết bị có khoảng cách thấp nhất so với người dùng

EEE2 Phương pháp hoặc hệ thống theo EEE1 trong đó việc phát hiện tín hiệu bao gồm tín hiệu từ đối tượng phát ra tiếng ồn bất kỳ hoặc điểm giám sát âm thanh mong muốn mà có mối liên hệ khoảng cách đã biết đến tập hợp các thiết bị.

EEE3. Phương pháp hoặc hệ thống theo EEE1 hoặc EEE2 với thứ tự các thiết bị bao gồm việc xem xét khoảng cách và tỷ lệ tín hiệu trên tiếng vọng của thiết bị đối với khoảng cách nguồn danh định.

EEE4. Phương pháp hoặc hệ thống theo EEE bất kỳ trong số các EEE từ EEE1 đến EEE3, với thứ tự xem xét độ gần được tổng quát hóa của thiết bị với người dùng, và tính tương hỗ gần đúng của việc này để ước lượng cải thiện tỷ lệ tín hiệu trên tiếng vọng hiệu quả nhất và sắp xếp các thiết bị theo cách này.

Các khía cạnh về một số phương án thực hiện được mô tả bao gồm hệ thống hoặc thiết bị được tạo cấu hình (ví dụ, được lập trình) để thực hiện một hoặc nhiều phương pháp được mô tả, và phương tiện đọc được bằng máy tính hữu hình (ví dụ, đĩa) lưu trữ mã để thực hiện một hoặc nhiều phương pháp hoặc các bước của phương pháp được mô tả. Ví dụ, hệ thống có thể là hoặc bao gồm bộ xử lý đa năng lập trình được, bộ xử lý tín hiệu số, hoặc bộ vi xử lý, phần mềm hoặc firmware được lập trình và/hoặc theo cách khác được tạo cấu hình để thực hiện thao tác bất kỳ trong số hàng loạt thao tác trên dữ liệu, bao gồm một hoặc nhiều phương pháp hoặc các bước của phương pháp được mô tả. Bộ xử lý đa năng như vậy có thể là hoặc bao gồm hệ thống máy tính gồm thiết bị đầu vào, bộ nhớ, và hệ thống phụ xử lý được lập trình (và/hoặc được tạo cấu hình theo cách

khác) để thực hiện một hoặc nhiều phương pháp (hoặc các bước của phương pháp) được mô tả để đáp lại dữ liệu đã được xác nhận như vậy.

Một số phương án được mô tả được thực hiện dưới dạng bộ xử lý tín hiệu số (DSP) có thể tạo cấu hình được (ví dụ, có thể lập trình) mà được tạo cấu hình (ví dụ, được lập trình và được tạo cấu hình khác) để thực hiện xử lý đã yêu cầu đối với (các) tín hiệu âm thanh, bao gồm việc thực hiện một hoặc nhiều phương pháp được mô tả. Ngoài ra, một số phương án (hoặc các phần tử của phương án) có thể được thực hiện dưới dạng bộ xử lý đa năng (ví dụ, máy tính cá nhân (personal computer - PC) hoặc hệ thống máy tính hoặc bộ vi xử lý khác, có thể bao gồm thiết bị đầu vào và bộ nhớ) được lập trình bằng phần mềm hoặc firmware và/hoặc được tạo cấu hình theo cách khác để thực hiện bất kỳ trong số nhiều thao tác gồm một hoặc nhiều phương pháp hoặc bước của phương pháp được mô tả. Ngoài ra, các phần tử của một số phương án được mô tả được thực hiện dưới dạng bộ xử lý đa năng hoặc DSP được tạo cấu hình (ví dụ, được lập trình) để thực hiện một hoặc nhiều phương pháp hoặc các bước của phương pháp được mô tả, và hệ thống cũng bao gồm các phần tử khác (ví dụ, một hoặc nhiều loa và/hoặc một hoặc nhiều micro). Bộ xử lý đa năng được tạo cấu hình để thực hiện một hoặc nhiều phương pháp hoặc bước của phương pháp được mô tả thường sẽ được ghép nối với thiết bị đầu vào (ví dụ, chuột và/hoặc bàn phím), bộ nhớ và thiết bị hiển thị.

Khía cạnh khác của một số phương án thực hiện được mô tả là phương tiện đọc được bằng máy tính (ví dụ, đĩa hoặc phương tiện lưu trữ hữu hình khác) lưu trữ mã để thực hiện (ví dụ, mã thực thi được để thực hiện) phương án bất kỳ trong số của một hoặc nhiều phương pháp hoặc các bước của phương pháp được mô tả.

Mặc dù các phương án và ứng dụng cụ thể đã được mô tả ở đây, nhưng đối với những người có hiểu biết trung bình trong lĩnh vực kỹ thuật này sẽ thấy rõ rằng có thể có nhiều biến thể đối với các phương án và ứng dụng được mô tả ở đây mà không nằm ngoài phạm vi của yêu cầu bảo hộ. Cần phải hiểu rằng mặc dù một số phương án thực hiện nhất định đã được thể hiện và mô tả, nhưng sáng chế không bị giới hạn ở các phương án cụ thể được mô tả và thể hiện hoặc các phương pháp cụ thể được mô tả.

YÊU CẦU BẢO HỘ

1. Phương pháp quản lý phiên âm thanh, phương pháp này bao gồm các bước:

nhận tín hiệu đầu ra từ mỗi micrô trong số nhiều micrô trong môi trường âm thanh, mỗi micrô trong số nhiều micrô nằm ở vị trí micrô của môi trường âm thanh, tín hiệu đầu ra bao gồm tín hiệu tương ứng với tiếng nói hiện tại của người;

xác định, dựa trên tín hiệu đầu ra, một hoặc nhiều khía cạnh của thông tin ngữ cảnh liên quan đến người, thông tin ngữ cảnh bao gồm ít nhất một vị trí hiện tại được ước lượng của người đó hoặc độ gần hiện tại được ước lượng của người đó với một hoặc nhiều vị trí micrô;

xác định thiết bị âm thanh được trang bị loa gần nhất tức là gần nhất với vị trí micrô gần nhất với vị trí hiện tại được ước lượng của người đó;

lựa chọn hai hoặc nhiều thiết bị âm thanh của môi trường âm thanh dựa, ít nhất một phần, vào một hoặc nhiều khía cạnh của thông tin ngữ cảnh, hai hoặc nhiều thiết bị âm thanh, mỗi thiết bị bao gồm ít nhất một loa và trong đó hai hoặc nhiều thiết bị âm thanh bao gồm thiết bị âm thanh được trang bị loa gần nhất;

xác định một hoặc nhiều kiểu thay đổi xử lý âm thanh để áp dụng cho dữ liệu âm thanh đang được kết xuất thành tín hiệu cấp cho loa cho hai hoặc nhiều thiết bị âm thanh, các thay đổi xử lý âm thanh có tác dụng làm tăng tỷ lệ lời nói trên tiếng vọng tại micrô gần nhất với vị trí hiện tại được ước lượng của người, trong đó tiếng vọng bao gồm ít nhất âm thanh nào đó được phát ra bởi hai hoặc nhiều thiết bị âm thanh và trong đó ít nhất một trong những thay đổi xử lý âm thanh đối với thiết bị âm thanh gần nhất khác với thay đổi xử lý âm thanh đối với thiết bị âm thanh thứ hai trong số ít nhất hai thiết bị âm thanh nói trên, và trong đó một hoặc nhiều kiểu thay đổi xử lý âm thanh làm giảm mức tái tạo loa cho thiết bị âm thanh gần nhất; và

khiến cho một hoặc nhiều kiểu thay đổi xử lý âm thanh được áp dụng.

2. Phương pháp theo điểm 1, trong đó một hoặc nhiều kiểu thay đổi xử lý âm thanh bao gồm bước sửa đổi phổ.

3. Phương pháp theo điểm 2, trong đó bước sửa đổi phổ bao gồm bước giảm mức dữ liệu âm thanh trong băng tần số từ 500 Hz đến 3 KHz.

4. Phương pháp theo điểm 1, trong đó một hoặc nhiều kiểu thay đổi xử lý âm thanh làm giảm mức tái tạo loa đối với các loa của hai hoặc nhiều thiết bị âm thanh.
5. Phương pháp theo điểm 1, trong đó bước lựa chọn hai hoặc nhiều thiết bị âm thanh của môi trường âm thanh bao gồm bước lựa chọn N thiết bị âm thanh được trang bị loa của môi trường âm thanh, N là số nguyên lớn hơn 2.
6. Phương pháp theo điểm 1, trong đó bước lựa chọn hai hoặc nhiều thiết bị âm thanh của môi trường âm thanh dựa, ít nhất một phần, vào vị trí hiện tại được ước lượng của người đó so với ít nhất một trong số vị trí micrô hoặc vị trí thiết bị âm thanh được trang bị loa.
7. Phương pháp theo điểm 1, trong đó một hoặc nhiều kiểu thay đổi xử lý âm thanh bao gồm bước thay đổi quy trình kết xuất để làm lệch việc kết xuất tín hiệu âm thanh khỏi vị trí hiện tại được ước lượng của người đó.
8. Phương pháp theo điểm 1, trong đó một hoặc nhiều kiểu thay đổi xử lý âm thanh bao gồm bước chèn ít nhất một khe vào ít nhất một băng tần số đã chọn của tín hiệu phát lại âm thanh.
9. Phương pháp theo điểm 1, trong đó một hoặc nhiều kiểu thay đổi xử lý âm thanh bao gồm nén dải động.
10. Phương pháp theo điểm 1, trong đó bước lựa chọn hai hoặc nhiều thiết bị âm thanh dựa, ít nhất một phần, vào ước lượng tỷ lệ tín hiệu trên tiếng vọng cho một hoặc nhiều vị trí micrô.
11. Phương pháp theo điểm 10, trong đó bước lựa chọn hai hoặc nhiều thiết bị âm thanh dựa, ít nhất một phần, vào bước xác định xem ước lượng tỷ lệ tín hiệu trên tiếng vọng có nhỏ hơn hoặc bằng ngưỡng tỷ lệ tín hiệu trên tiếng vọng hay không.
12. Phương pháp theo điểm 10, trong đó bước xác định một hoặc nhiều kiểu thay đổi xử lý âm thanh được dựa trên việc tối ưu hóa hàm chi phí dựa, ít nhất một phần, vào ước lượng tỷ lệ tín hiệu trên tiếng vọng.

13. Phương pháp theo điểm 12, trong đó hàm chi phí dựa, ít nhất một phần, vào hiệu suất kết xuất.

14. Phương pháp theo điểm 1, trong đó bước lựa chọn hai hoặc nhiều thiết bị âm thanh dựa, ít nhất một phần, vào ước lượng độ gần.

15. Phương pháp theo điểm 1, phương pháp này còn bao gồm các bước:

xác định nhiều đặc tính âm thanh hiện tại từ tín hiệu đầu ra của mỗi micrô;

áp dụng bộ phân loại cho nhiều đặc tính âm thanh hiện tại, trong đó bước áp dụng bộ phân loại bao gồm bước áp dụng mô hình được huấn luyện trên các đặc tính âm thanh được xác định từ trước thu được từ nhiều tiếng nói trước đó của một người trong nhiều vùng người dùng trong môi trường; và

trong đó bước xác định một hoặc nhiều khía cạnh thông tin ngữ cảnh liên quan đến người bao gồm bước xác định, dựa ít nhất một phần vào đầu ra từ bộ phân loại, ước lượng về vùng người dùng mà người đó hiện đang ở.

16. Phương pháp theo điểm 15, trong đó bước ước lượng về vùng người dùng được xác định mà không cần tham chiếu đến các vị trí hình học của nhiều micrô.

17. Phương pháp theo điểm 15, trong đó tiếng nói hiện tại và các tiếng nói trước đó bao gồm các tiếng nói từ đánh thức.

18. Phương pháp theo điểm 1, phương pháp này còn bao gồm bước lựa chọn ít nhất một micrô theo một hoặc nhiều khía cạnh của thông tin ngữ cảnh.

19. Phương pháp theo điểm 1, trong đó một hoặc nhiều micrô nằm trong nhiều thiết bị âm thanh của môi trường âm thanh.

20. Phương pháp theo điểm 1, trong đó một hoặc nhiều micrô nằm trong một thiết bị âm thanh của môi trường âm thanh.

21. Phương pháp theo điểm 1, trong đó ít nhất một trong số một hoặc nhiều vị trí micrô tương ứng với nhiều micrô của một thiết bị âm thanh.

22. Thiết bị quản lý phiên âm thanh được tạo cấu hình để thực hiện phương pháp theo điểm 1.

23. Hệ thống quản lý phiên âm thanh được tạo cấu hình để thực hiện phương pháp theo 1.

24. Phương tiện bất biến có phần mềm được lưu trữ trên đó, phần mềm bao gồm các các lệnh để điều khiển một hoặc nhiều thiết bị để thực hiện phương pháp theo điểm 1.

1 / 25

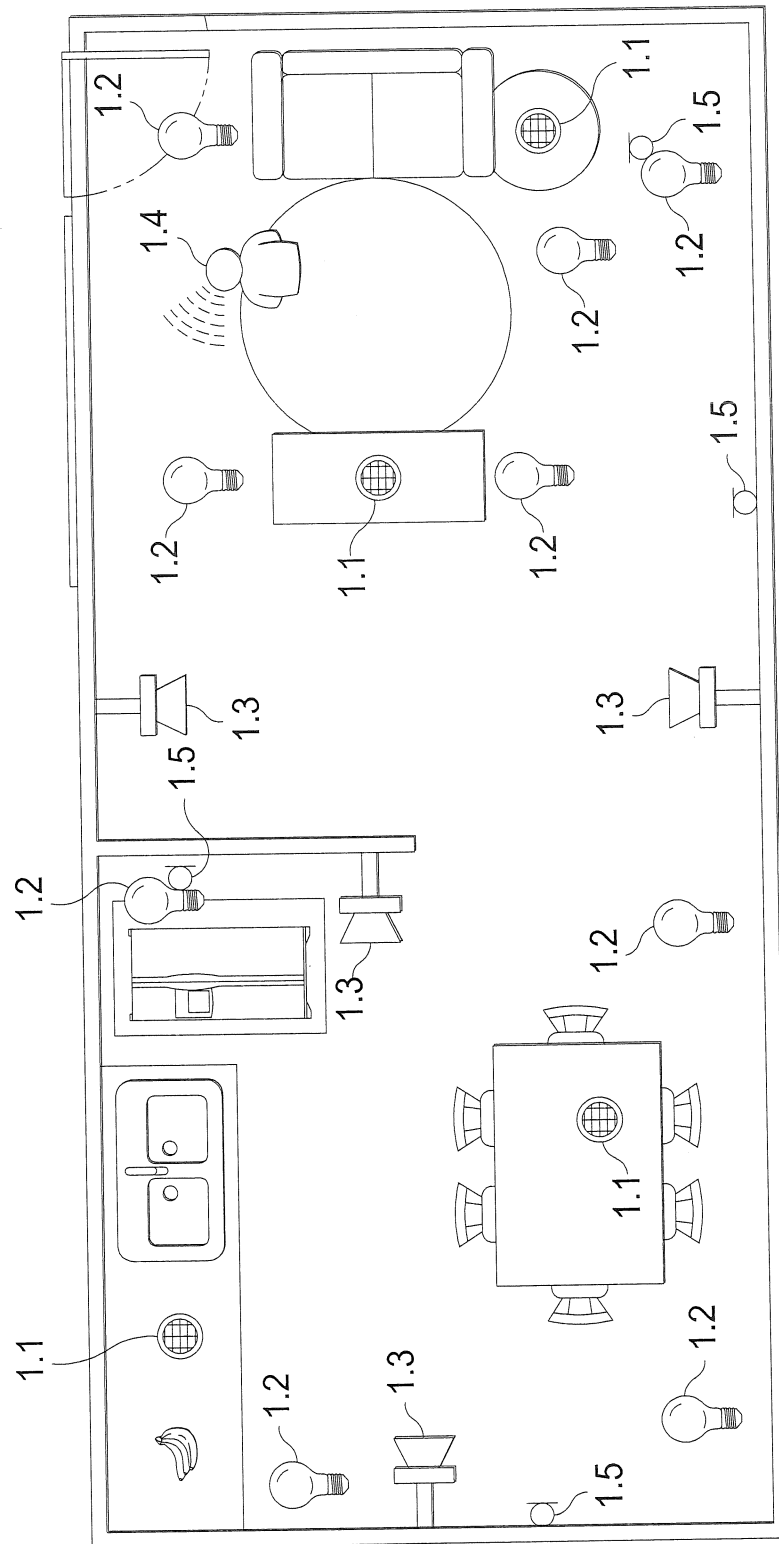


Fig. 1A

2 / 25

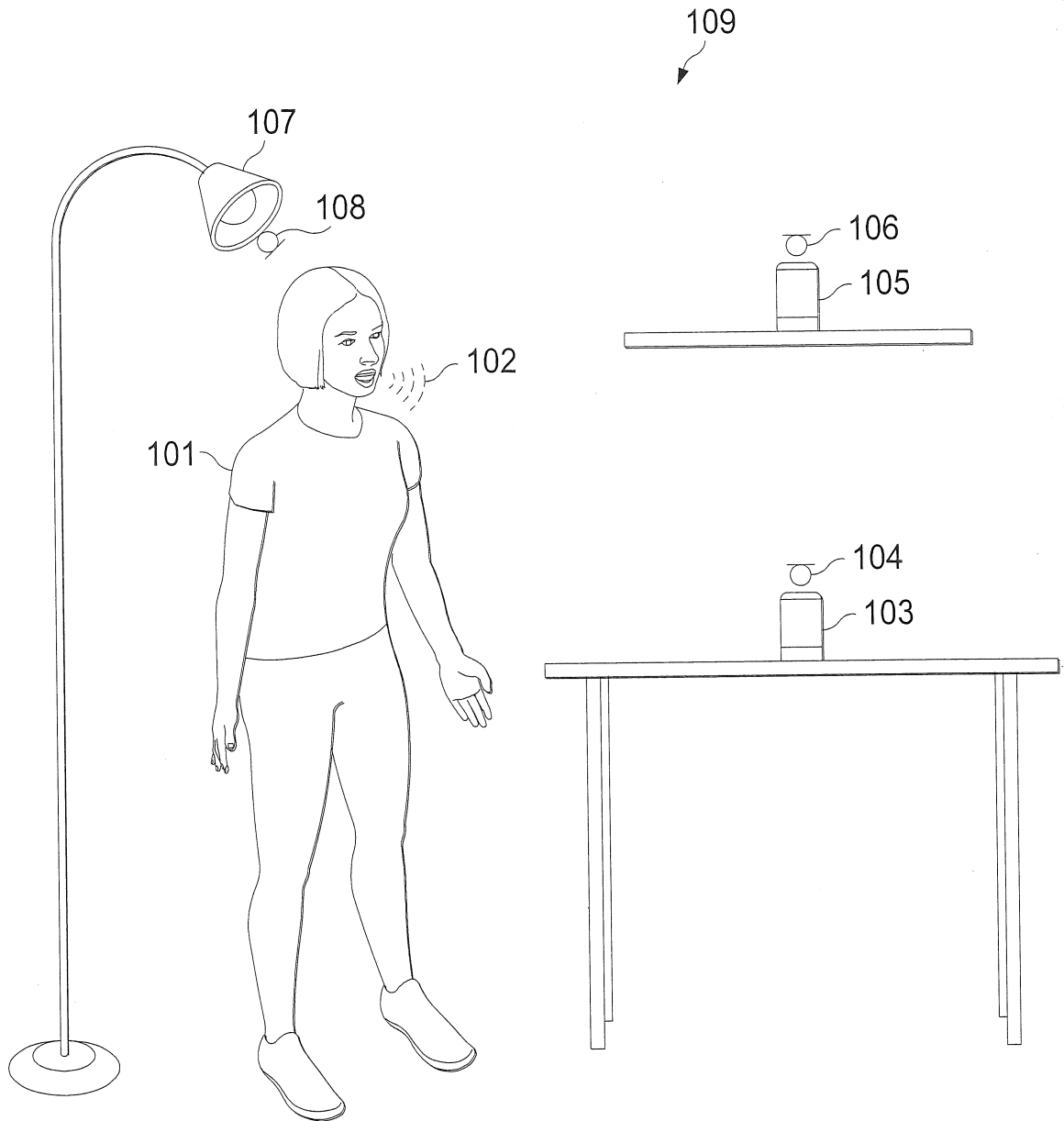


Fig.1B

3 / 25

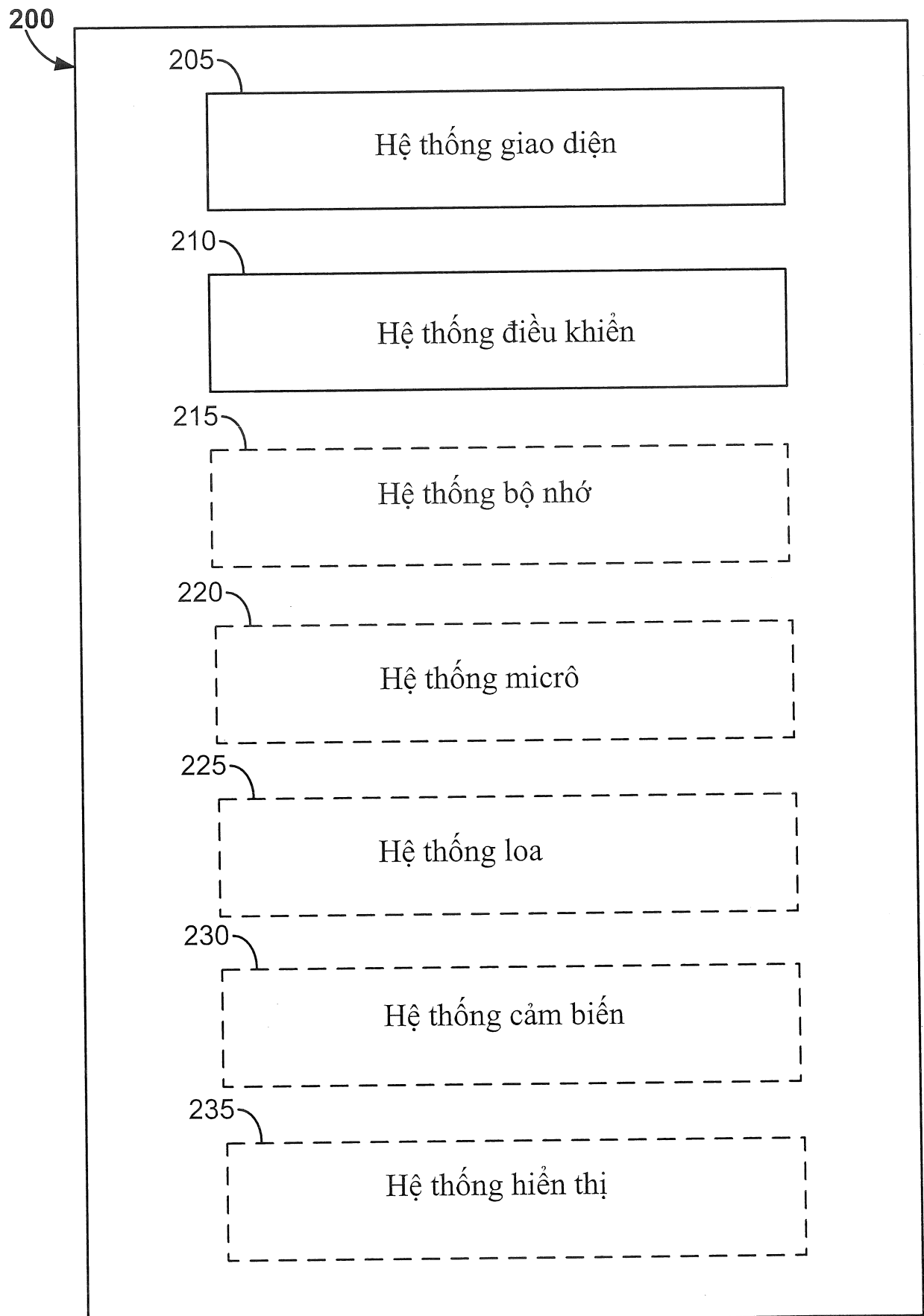
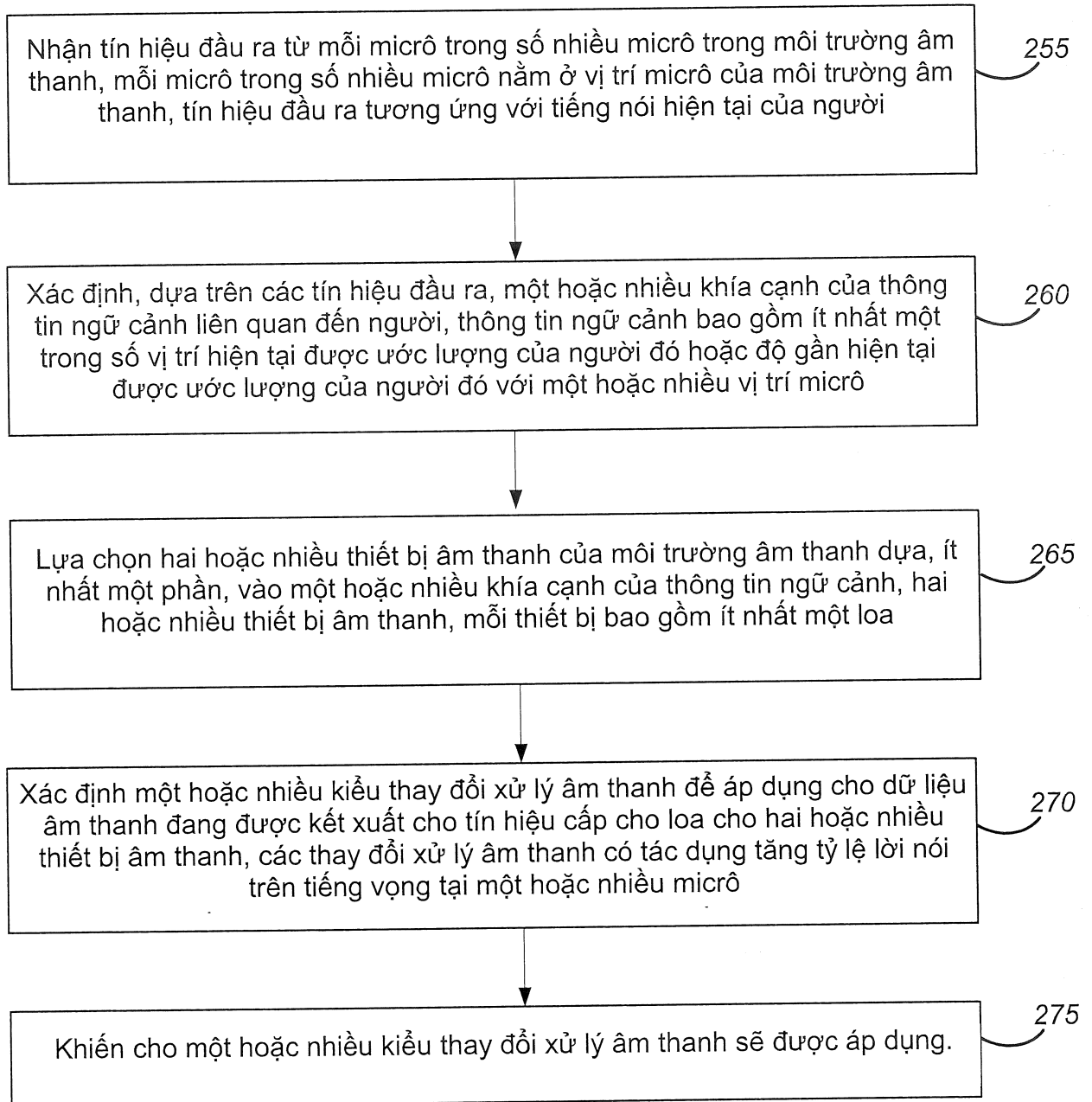


Fig.2A

4 / 25



250

Fig.2B

5 / 25

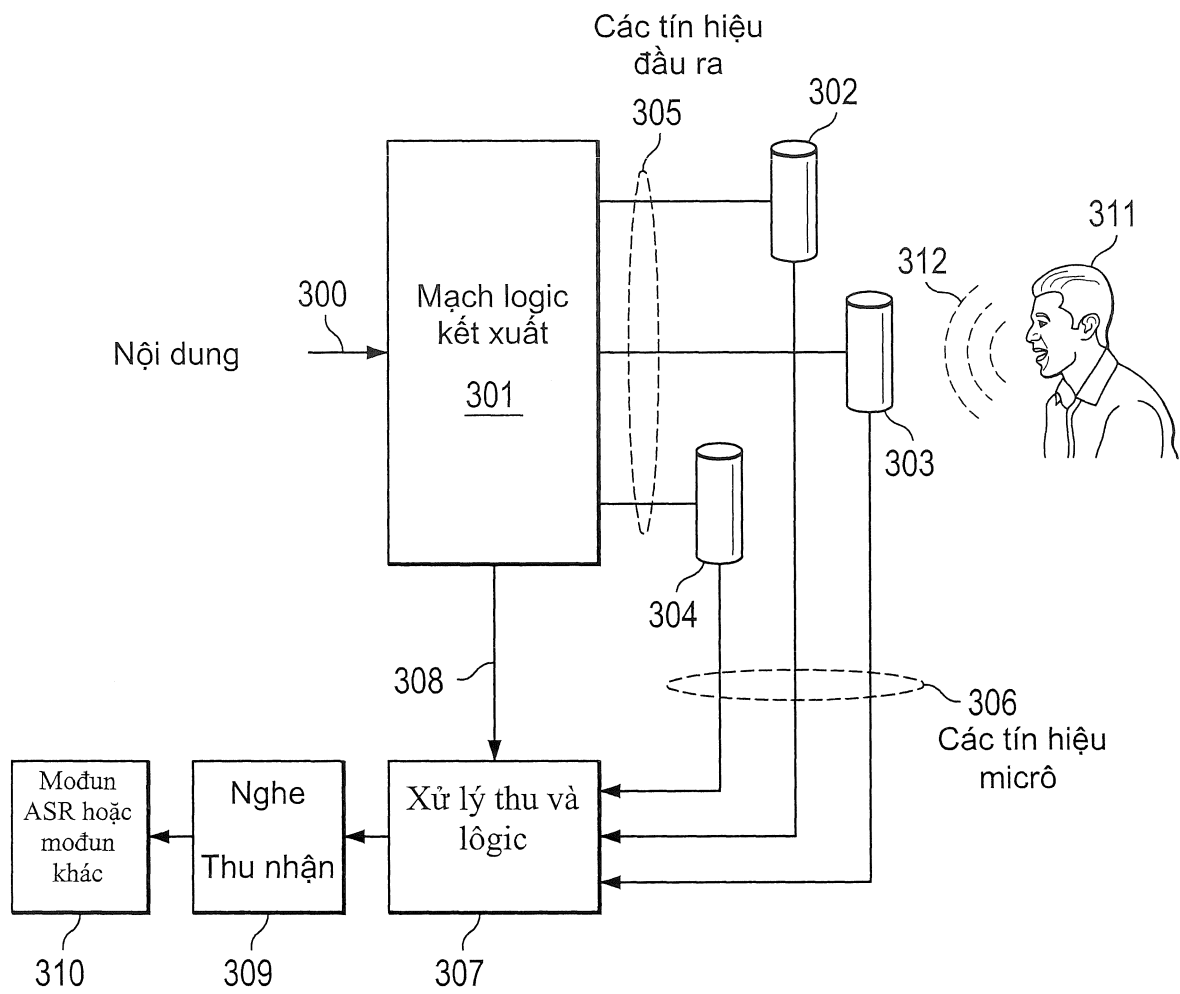


Fig.3A

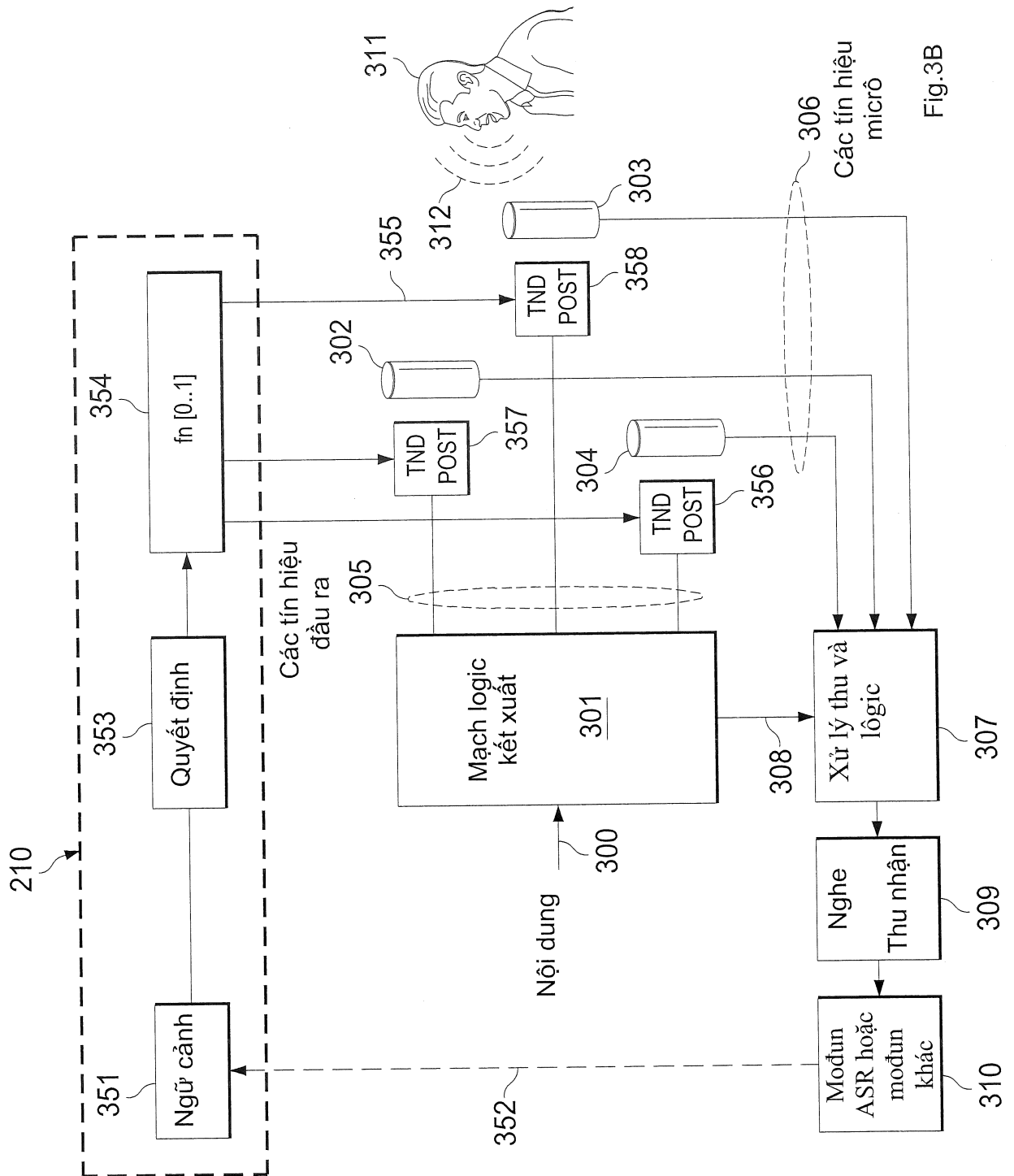


Fig.3B

7 / 25

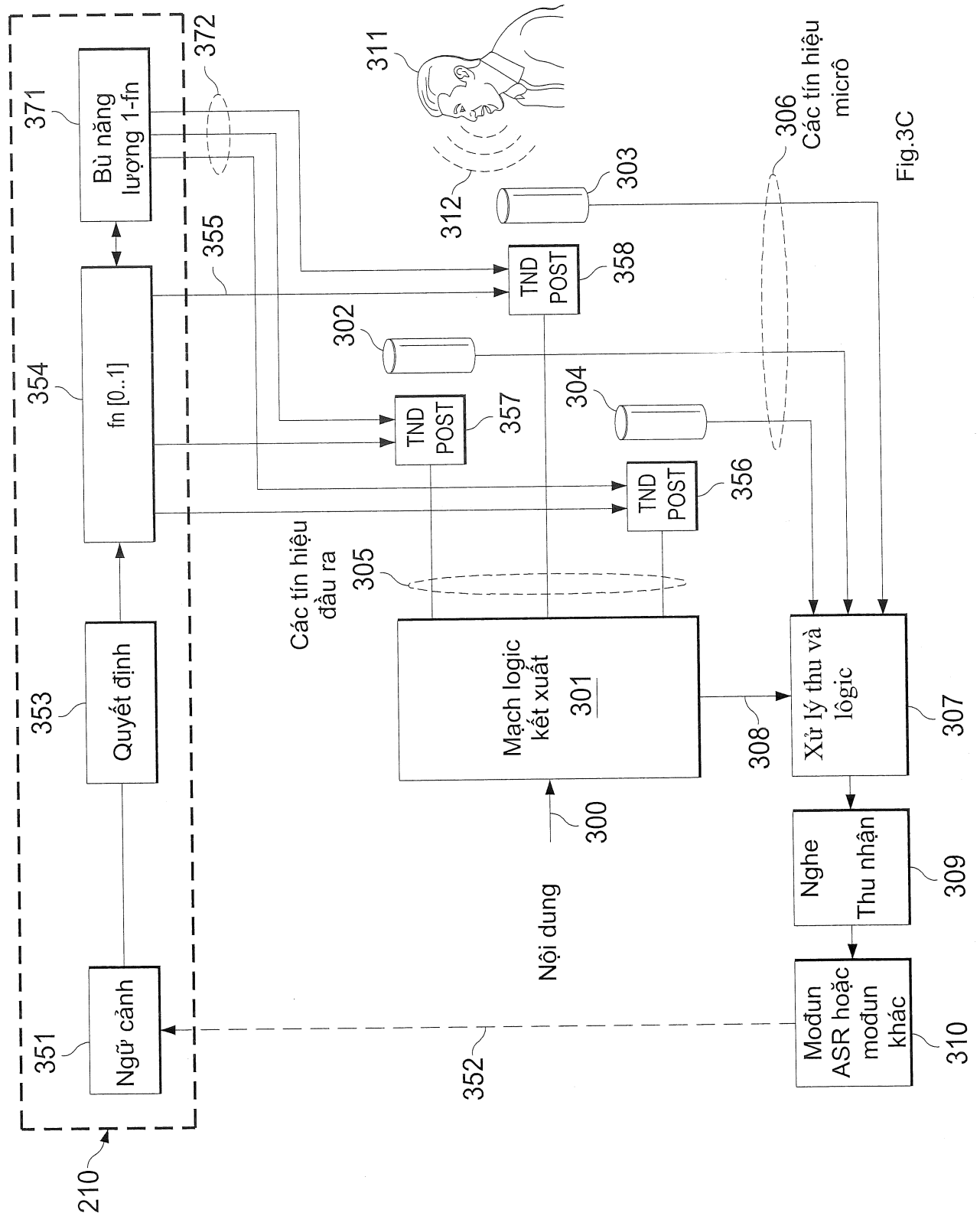


Fig.3C

8 / 25

Fig.4

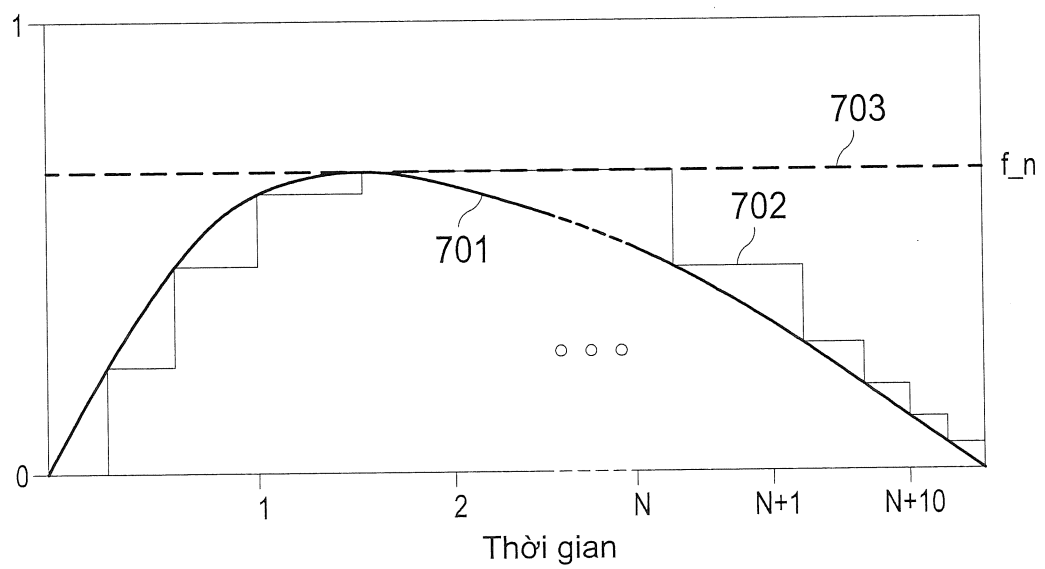
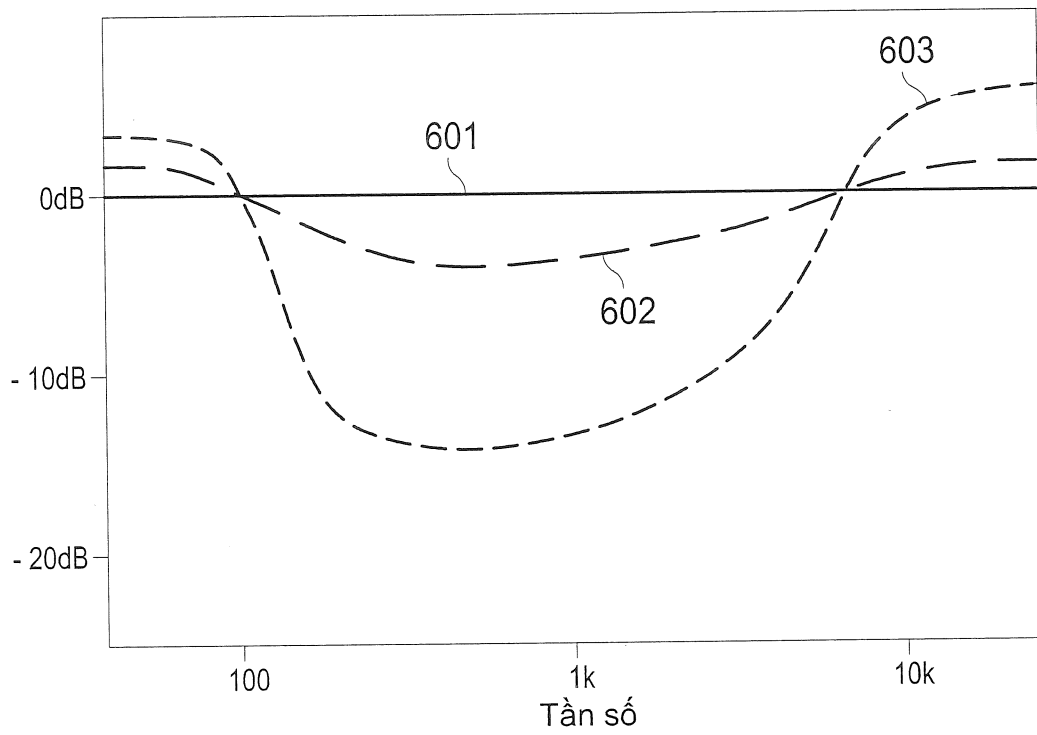


Fig.5

9 / 25

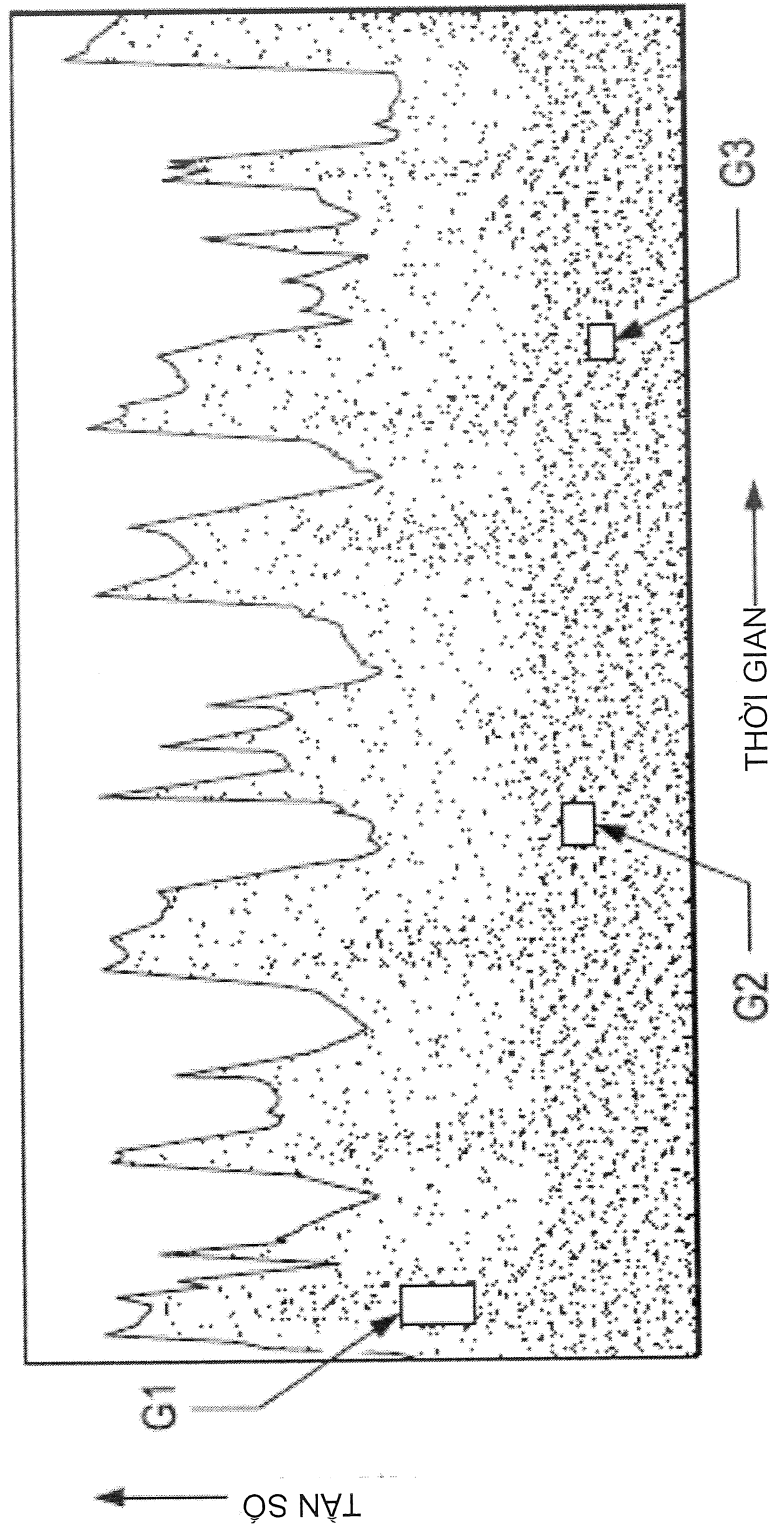


Fig.6

10 / 25

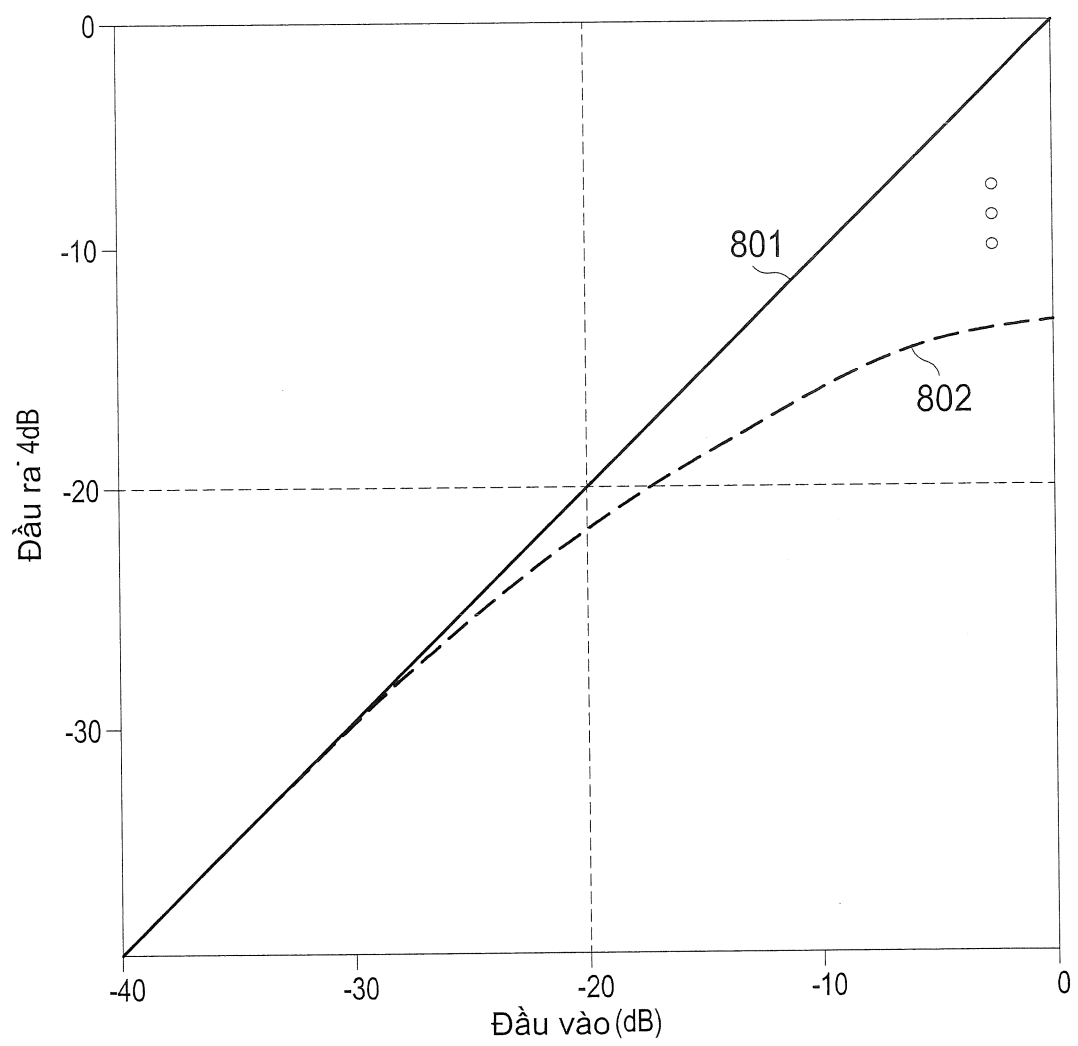


Fig.7

11 / 25

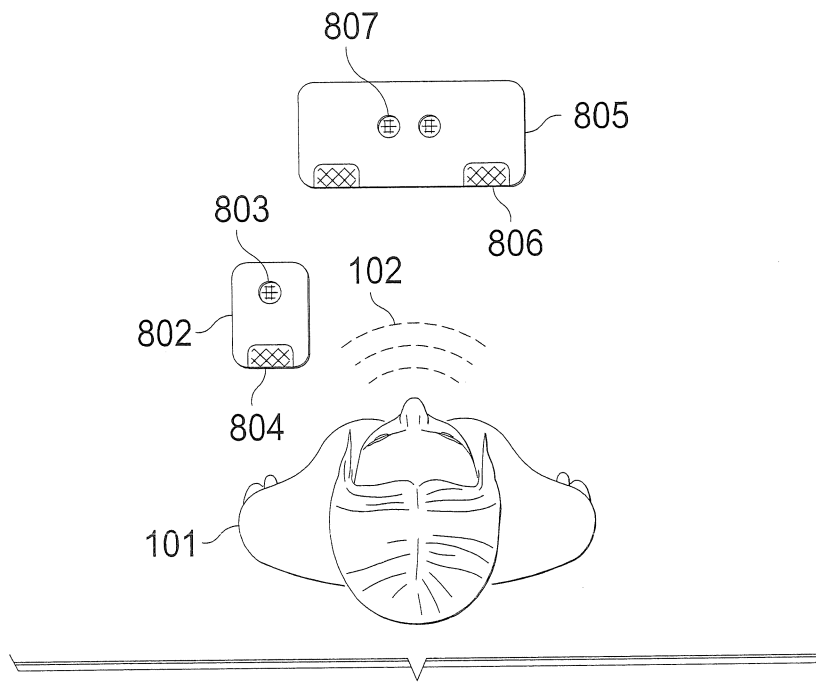


Fig.8

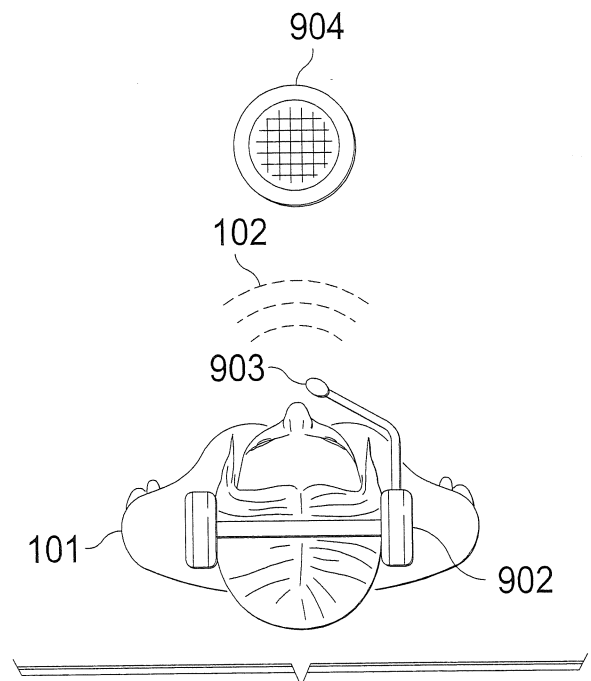


Fig.9

12 / 25

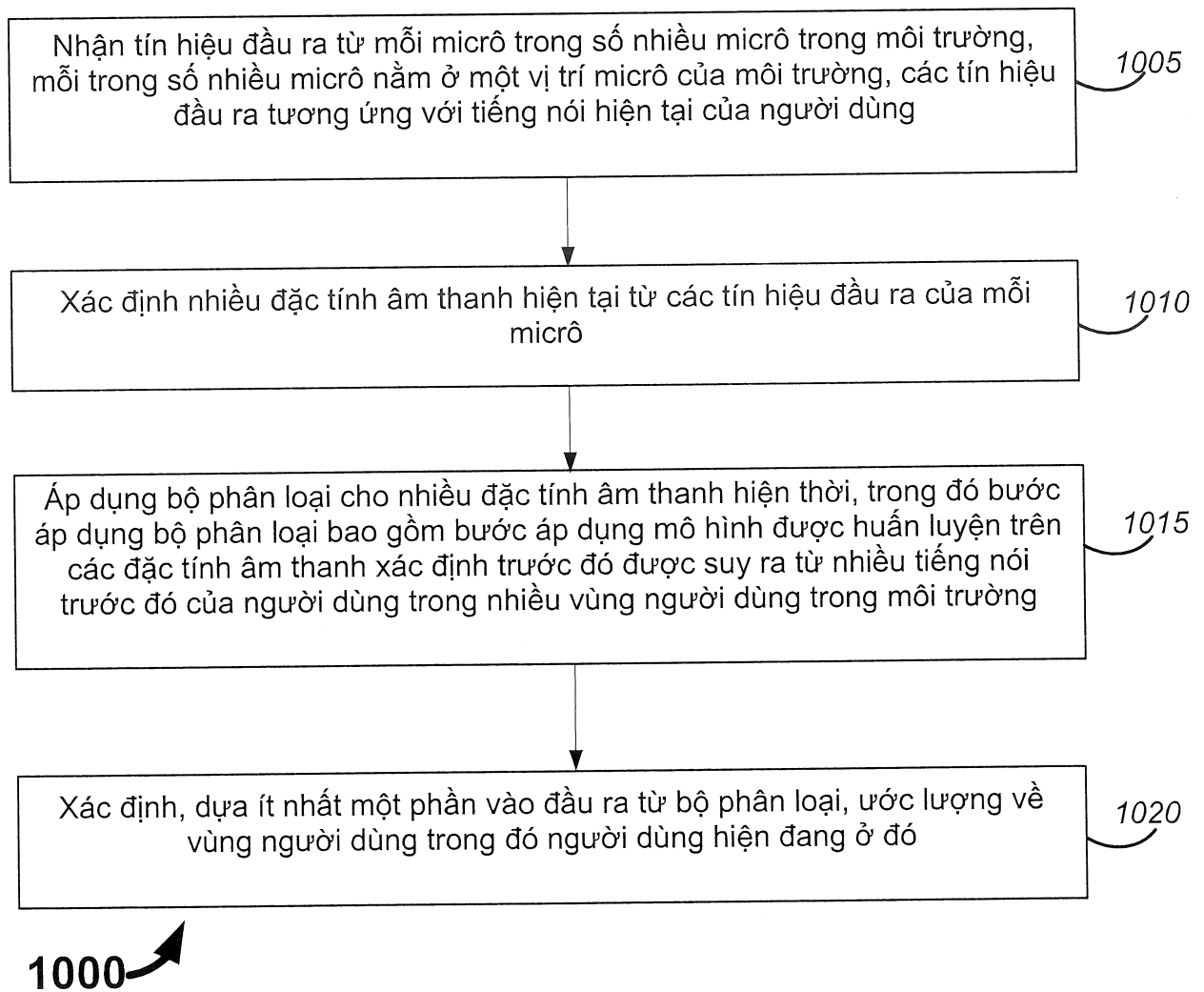


Fig.10

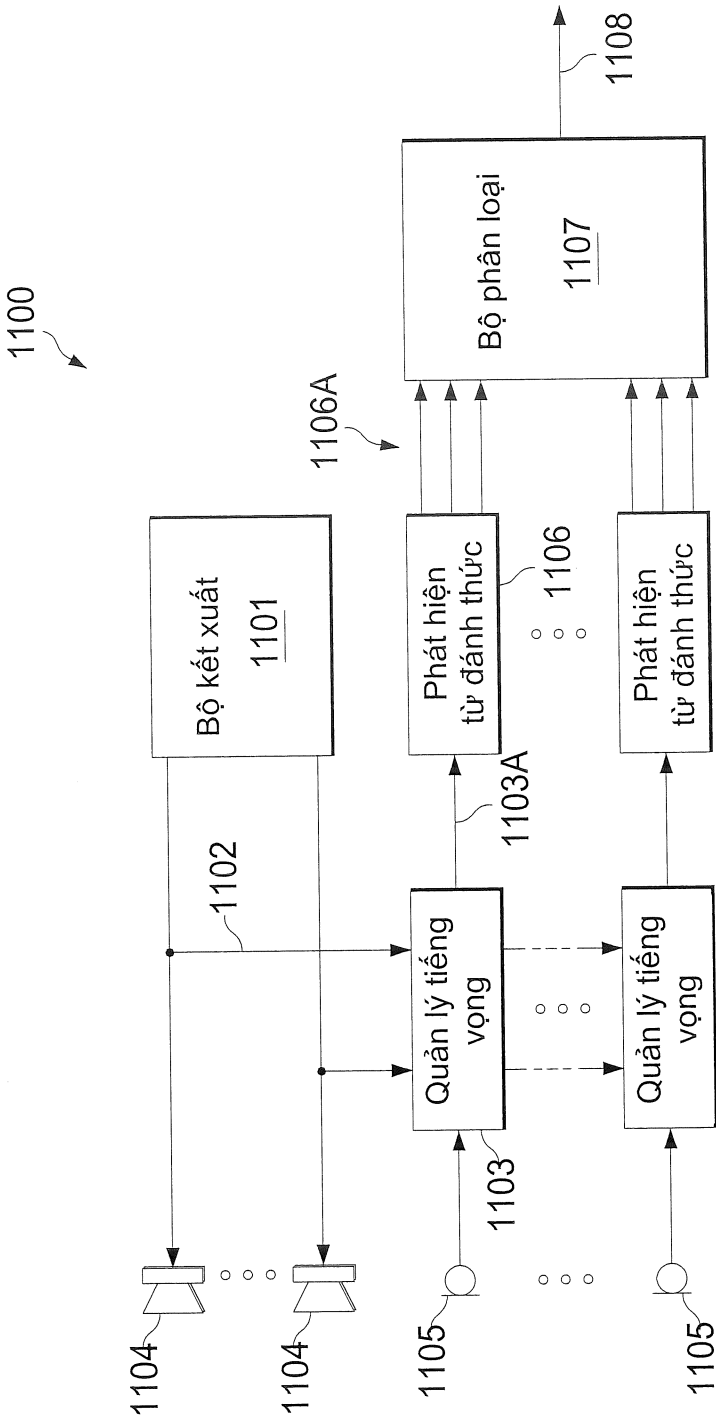


Fig.11

14 / 25

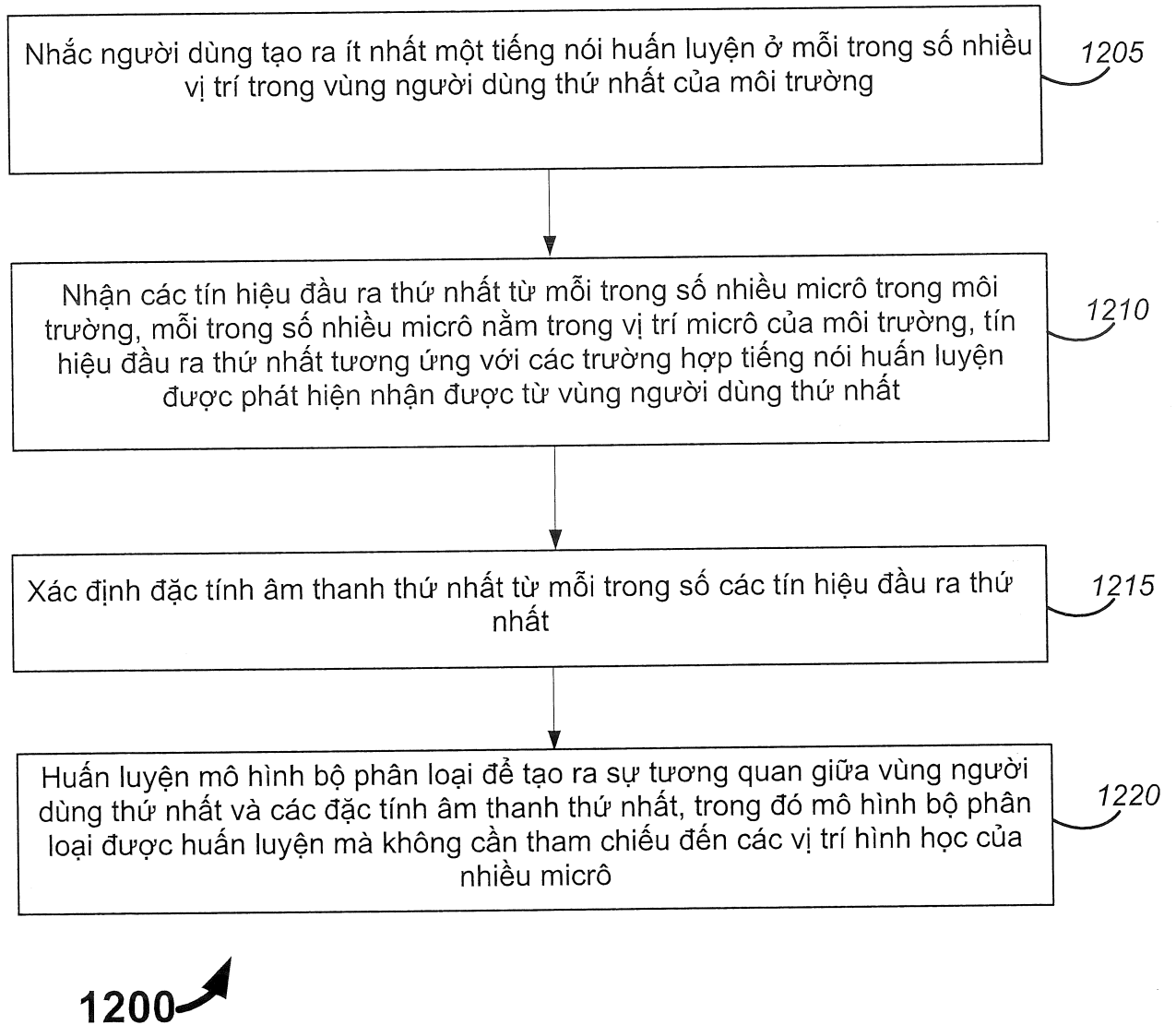


Fig.12

15 / 25

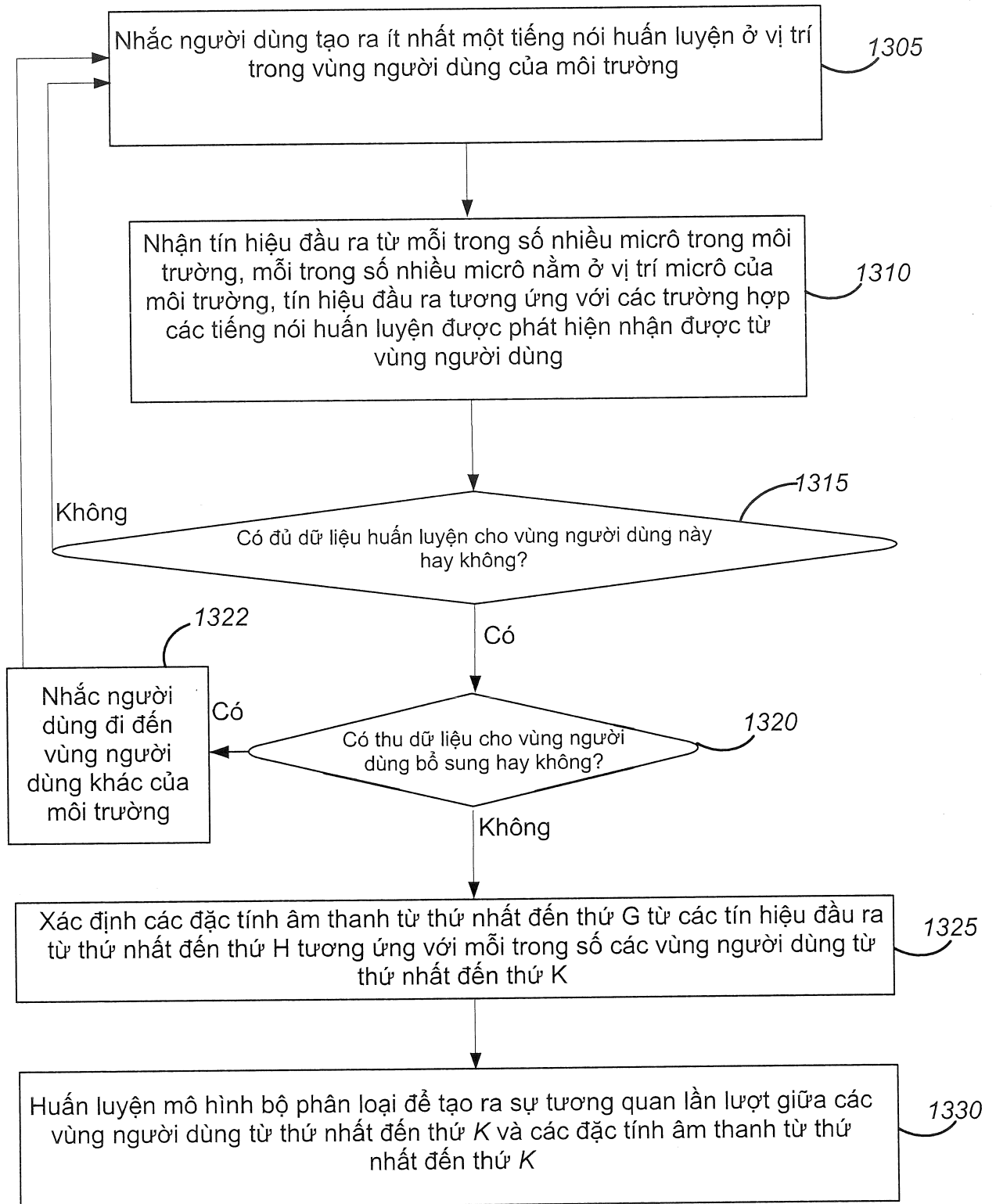


Fig.13

1300

16 / 25

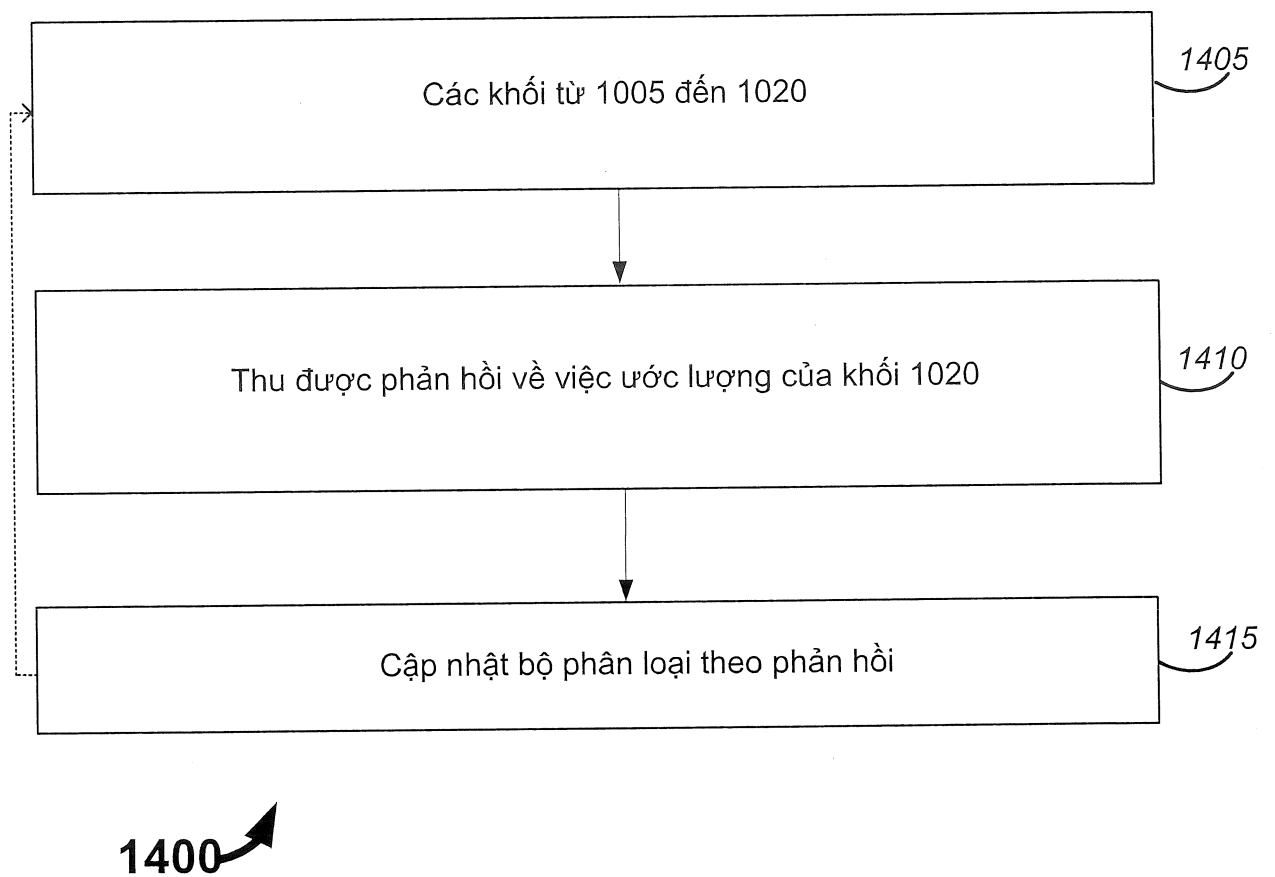
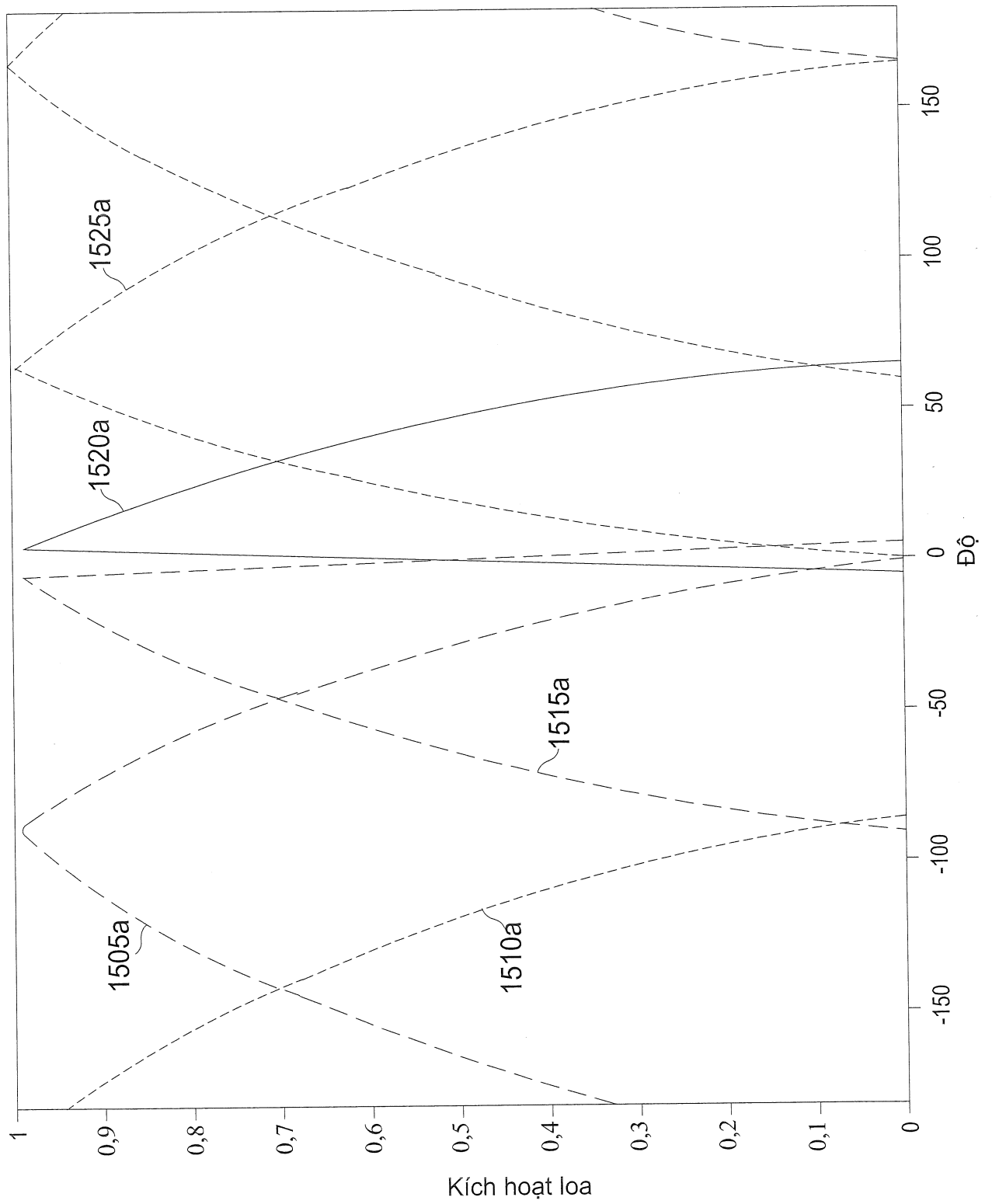


Fig.14

17 / 25



18 / 25

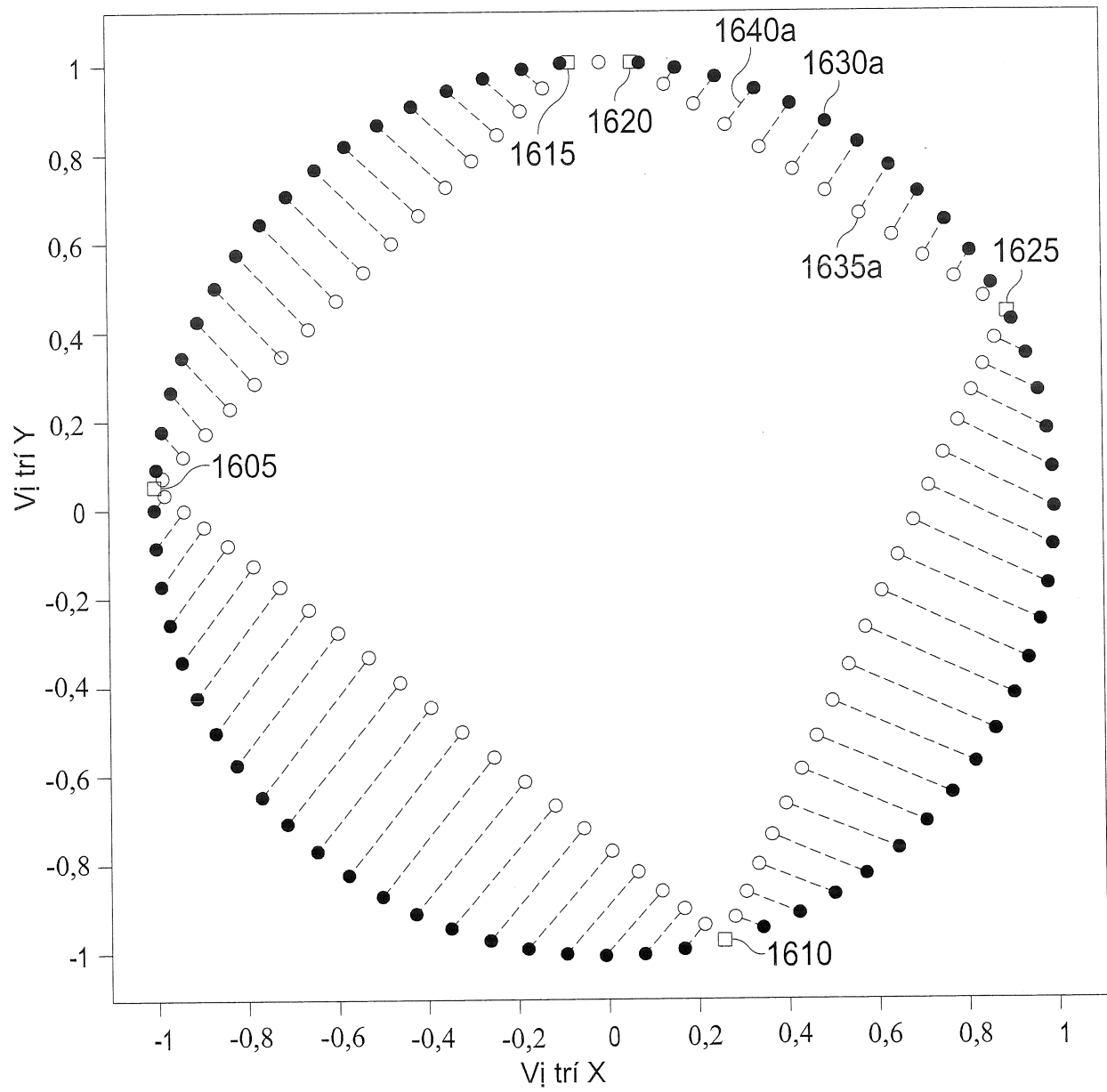


Fig.16

19 / 25

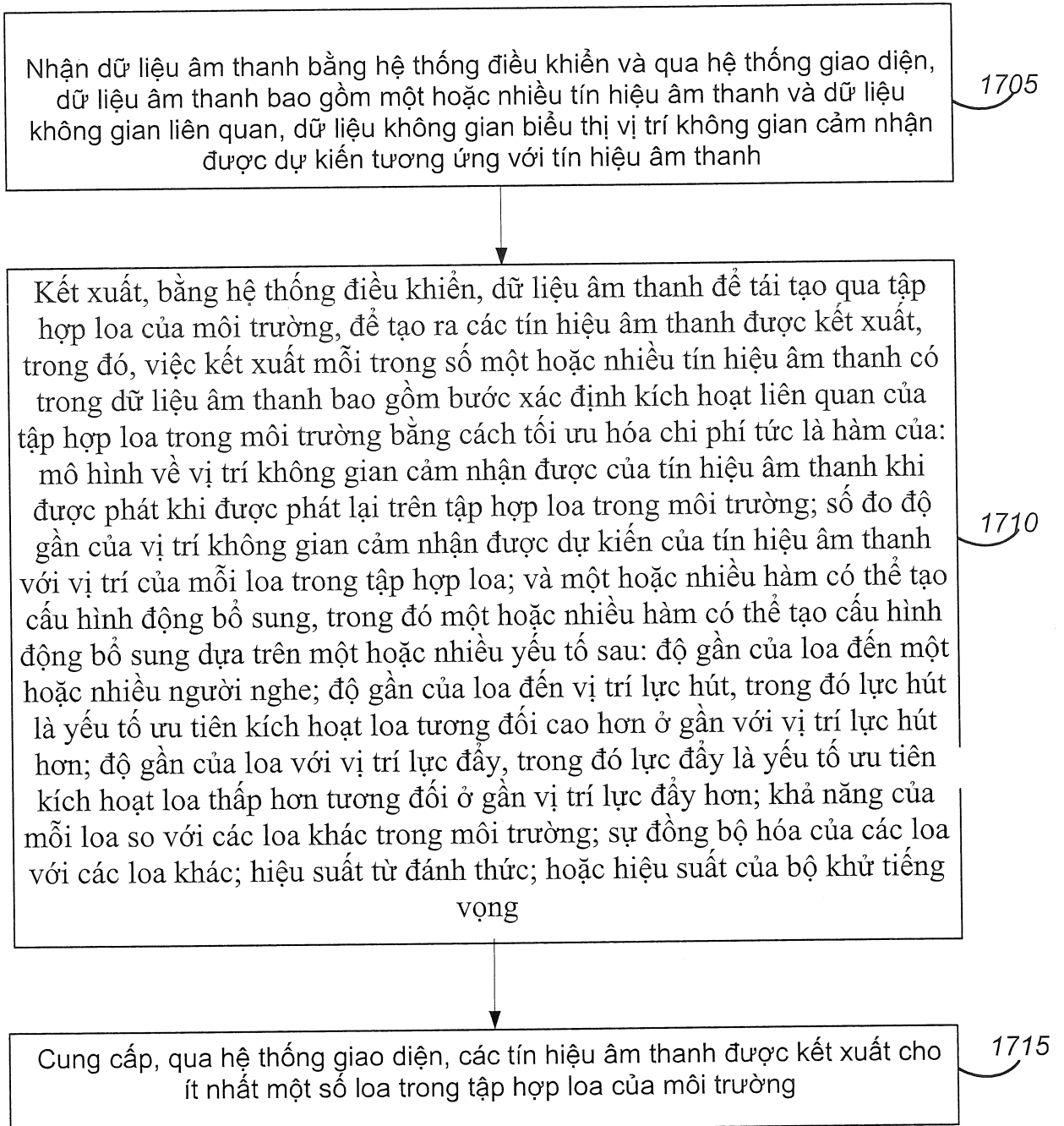


Fig.17

1700

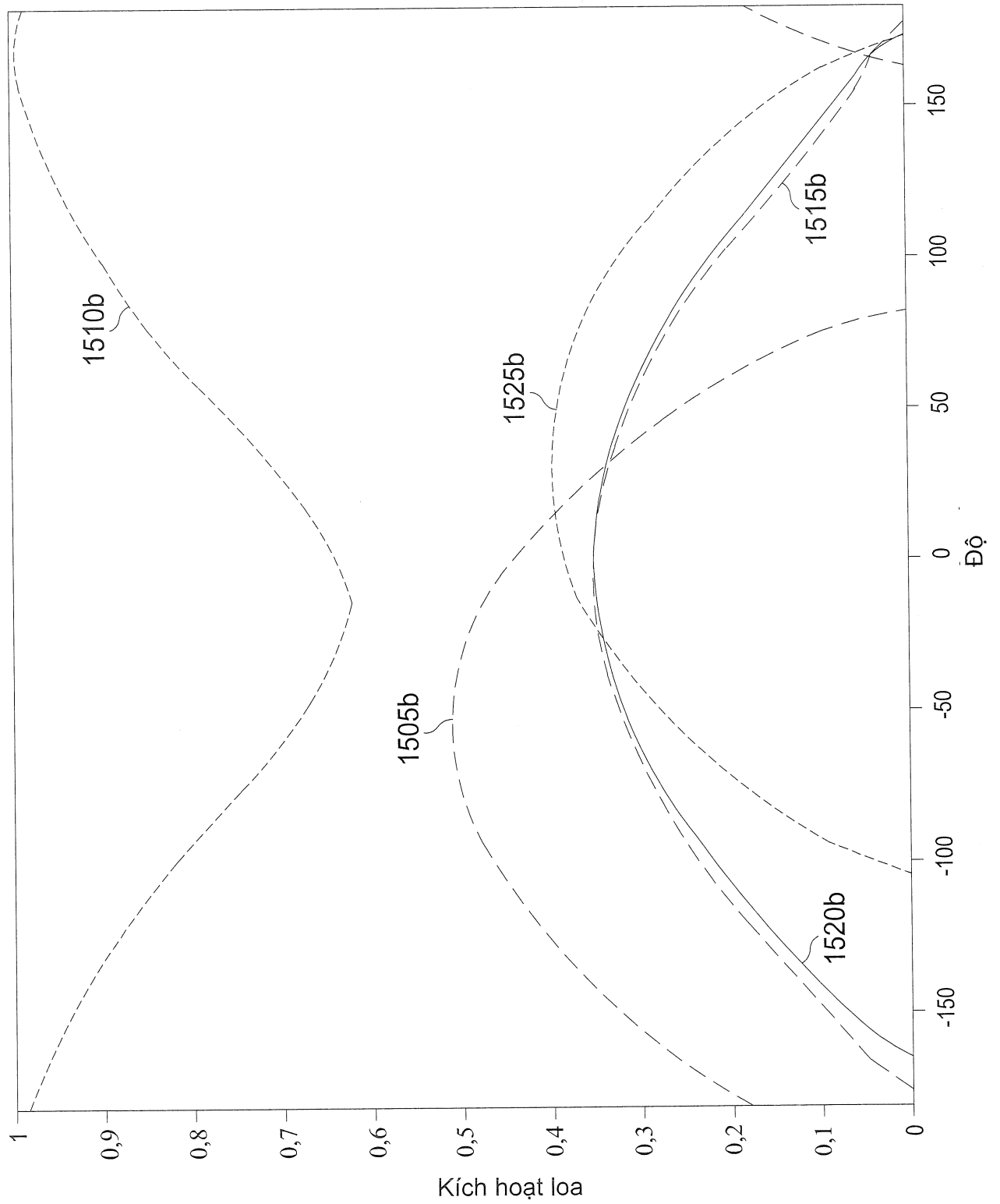


Fig.18

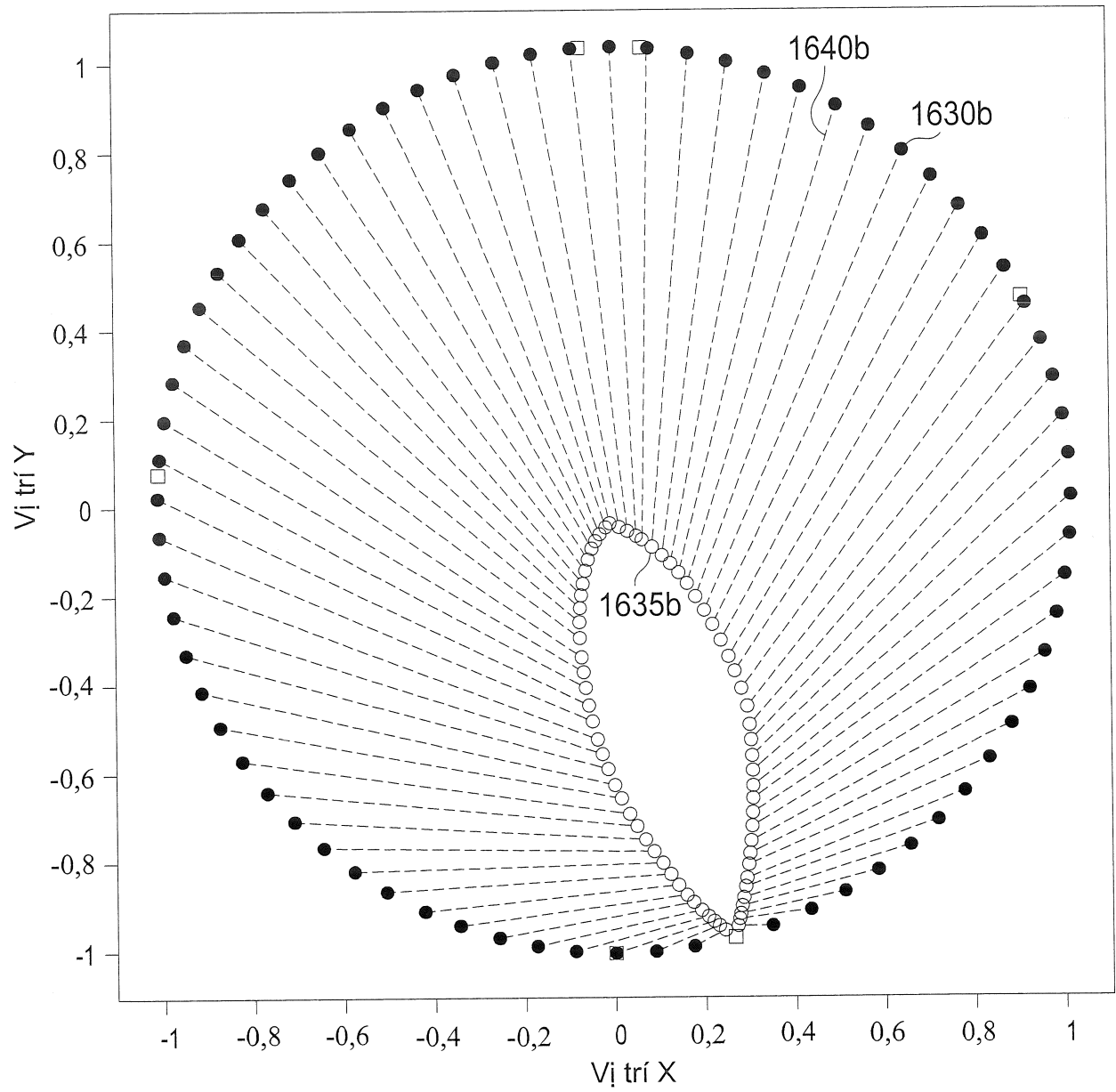
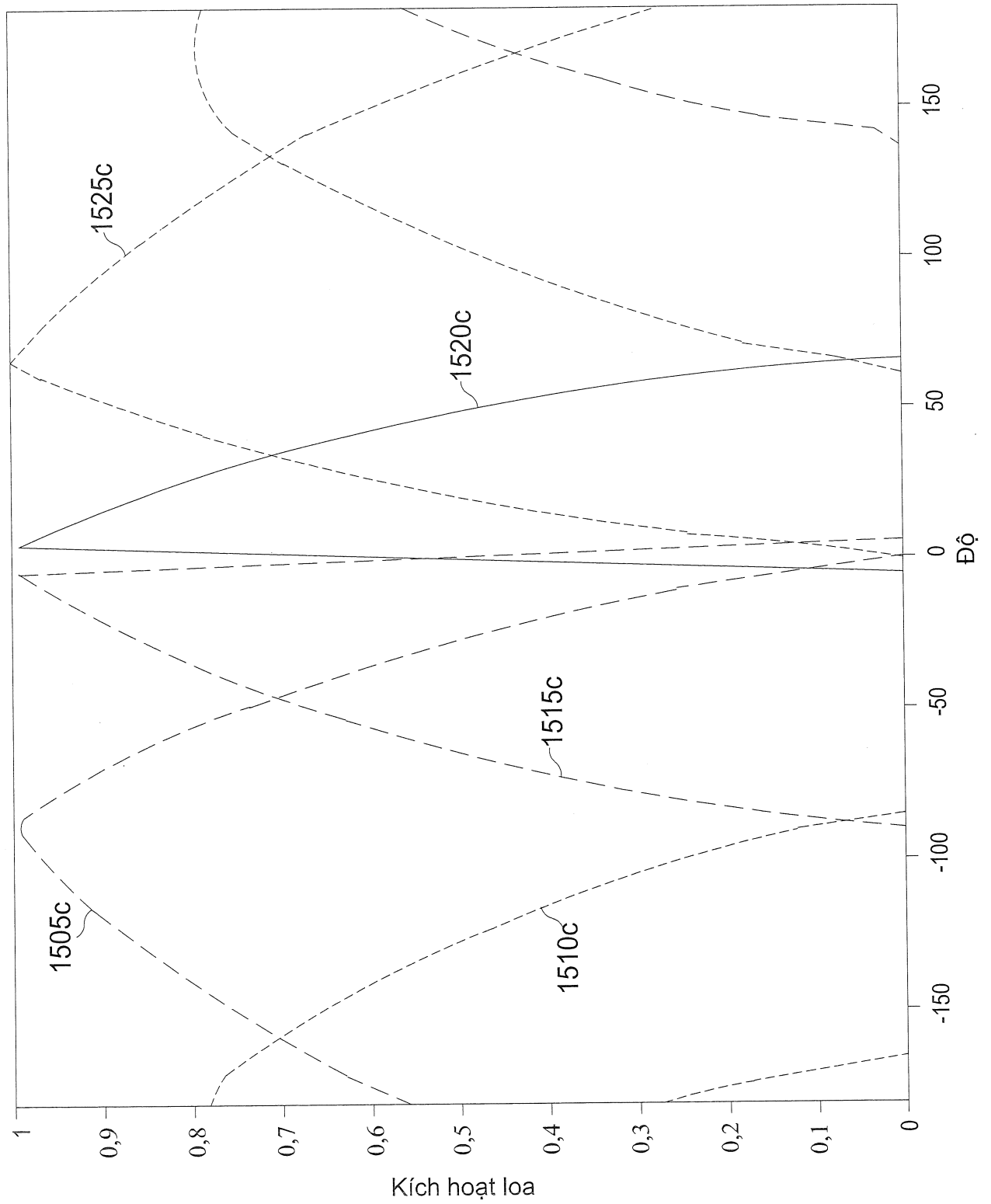


Fig.19

22 / 25



23 / 25

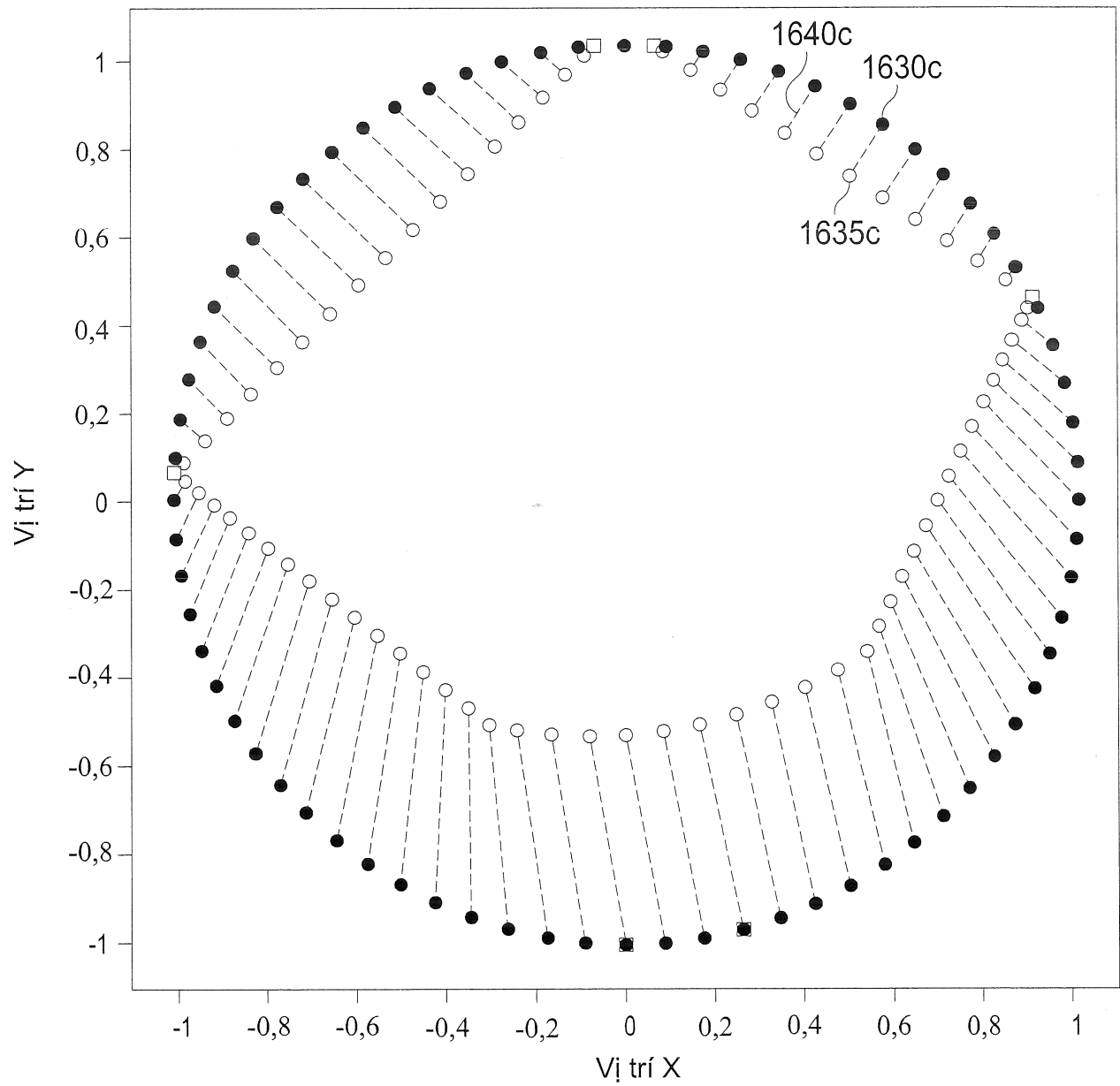


Fig.21

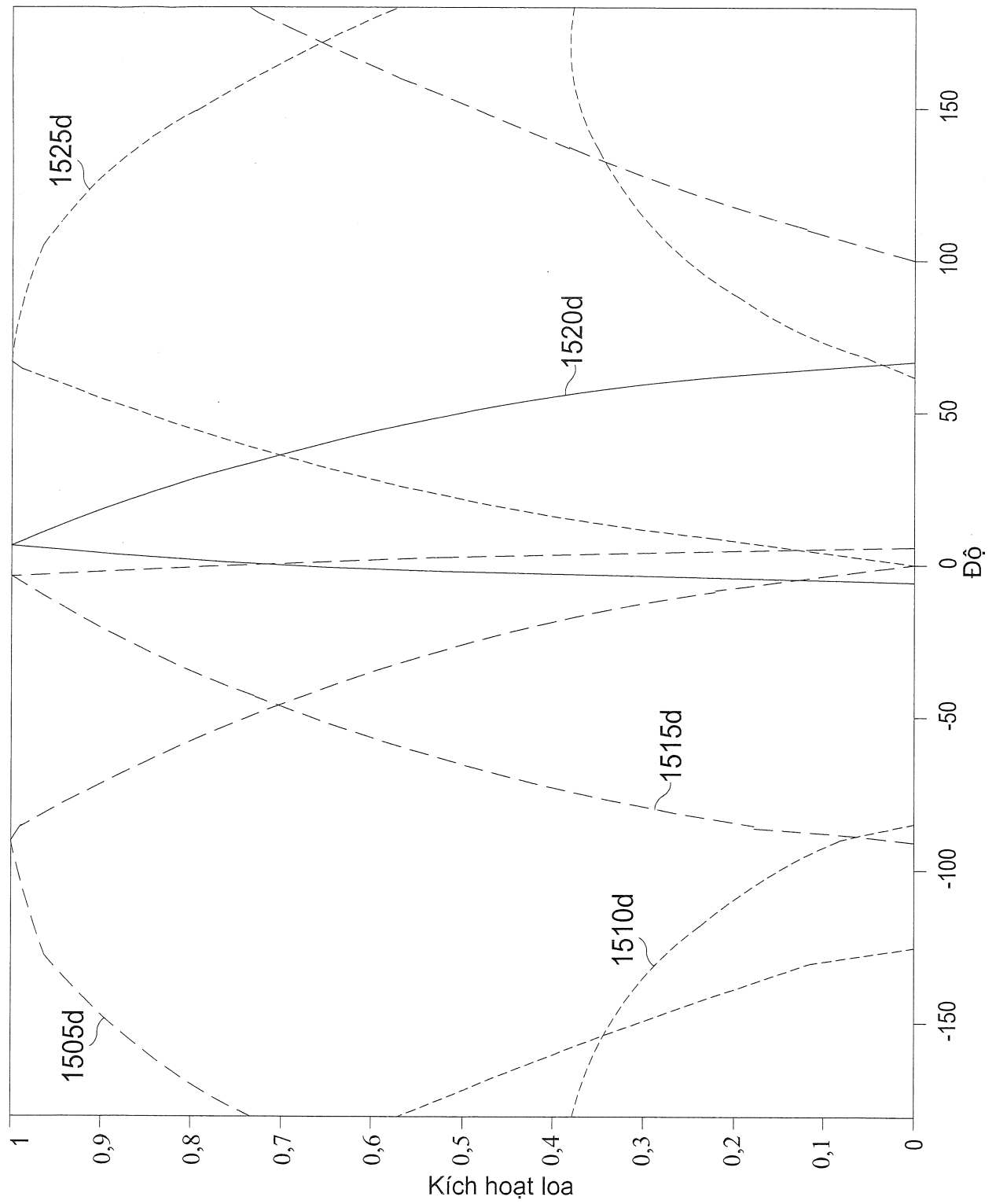


Fig.22

25 / 25

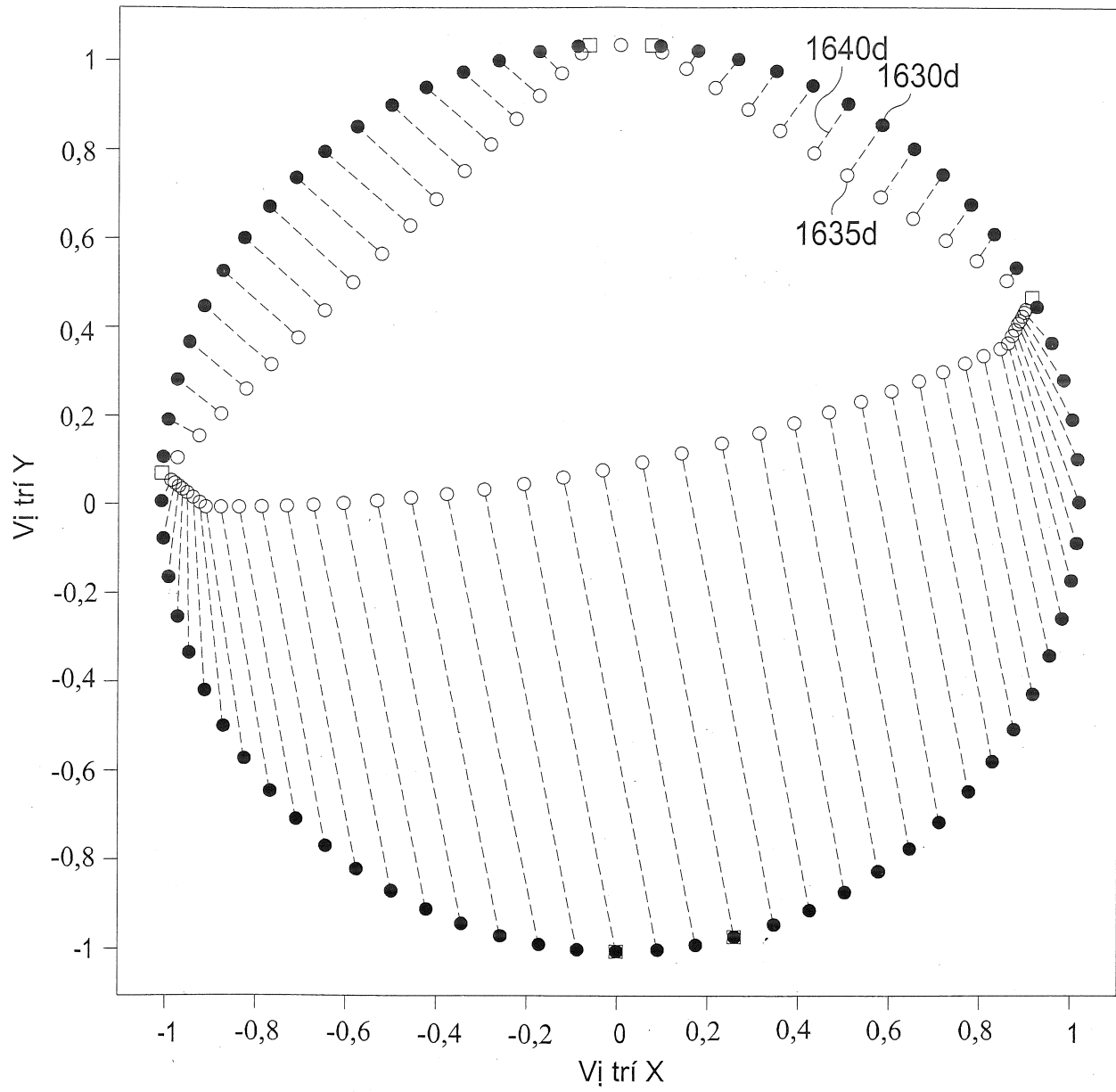


Fig.23