

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2015 年 3 月 5 日(05.03.2015)



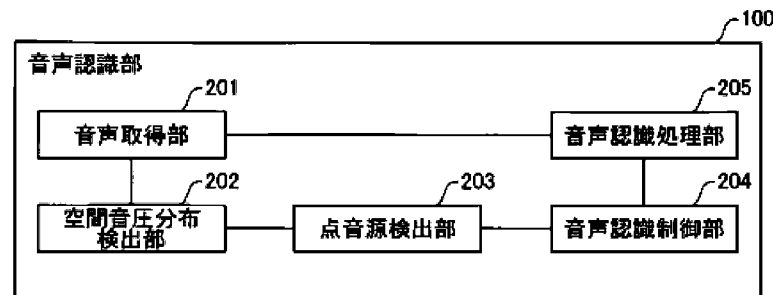
(10) 国際公開番号

WO 2015/029296 A1

- (51) 国際特許分類:
G10L 15/28 (2013.01) G10L 25/51 (2013.01)
- (21) 国際出願番号: PCT/JP2014/003459
- (22) 国際出願日: 2014 年 6 月 30 日(30.06.2014)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
61/871,544 2013 年 8 月 29 日(29.08.2013) US
61/910,193 2013 年 11 月 29 日(29.11.2013) US
- (71) 出願人: パナソニック インテレクチュアル
プロパティ コーポレーション オブ アメリ
カ (PANASONIC INTELLECTUAL PROPERTY
CORPORATION OF AMERICA) [US/US]; 90503 カ
リフォルニア州、トーランス、スイート 20
0、マリナー アベニュー 20000 Cali
fornia (US).
- (72) 発明者: 西川 剛樹(NISHIKAWA, Tsuyoki).
- (74) 代理人: 徳田 佳昭, 外(TOKUDA, Yoshiaki et al.);
〒5718501 大阪府門真市大字門真 1006 番地
パナソニック株式会社内 Osaka (JP).
- (81) 指定国 (表示のない限り、全ての種類の国内保
護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA,
BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN,
CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES,
FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN,
IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR,
LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX,
MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH,
PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK,
SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA,
UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) 指定国 (表示のない限り、全ての種類の広域保
護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW,
MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシ
ア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ
(AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR,
GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT,
NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI
(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML,
MR, NE, SN, TD, TG).
- 添付公開書類:
— 国際調査報告 (条約第 21 条(3))
— 補正された請求の範囲 (条約第 19 条(1))

(54) Title: SPEECH RECOGNITION METHOD AND SPEECH RECOGNITION DEVICE

(54) 発明の名称: 音声認識方法及び音声認識装置



- 100 Speech recognition unit
201 Speech acquisition unit
202 Spatial sound pressure distribution detection unit
203 Point sound source detection unit
204 Speech recognition control unit
205 Speech recognition processing unit

(57) Abstract: A speech recognition method for recognizing speech for operating equipment includes: a speech information acquisition step for acquiring speech information from a microphone disposed in a prescribed space; a spatial sound pressure distribution detection step for detecting spatial sound pressure distribution showing the distribution of sound pressure in the space on the basis of the speech information that has been acquired; a point sound source detection step for detecting a point sound source within the space on the basis of the spatial sound pressure distribution that has been detected; and a speech recognition control step in which a determination is made to carry out speech recognition processing for the speech information that has been acquired when a point sound source has been detected.

(57) 要約: 機器を動作させるための音声認識する音声認識方法は、所定の空間に配置されたマイクロフォンから音声情報を取得する音声情報取得ステップと、取得された音声情報に基づいて、空間内における音圧の分布を示す空間音圧分布を検出する空間音圧分布検出ステップと、検出された空間音圧分布に基づいて、空間内における点音源を検出する点音源検出ステップと、点音源が検出された場合、取得された音声情報に対して音声認識処理を行うと判断する音声認識制御ステップとを含む。

WO 2015/029296 A1

明 細 書

発明の名称： 音声認識方法及び音声認識装置

技術分野

[0001] 本開示は、ユーザが発話した音声进行認識する音声認識方法及び音声認識装置に関するものである。

背景技術

[0002] 音声認識の誤動作を防止する技術が検討されている。

[0003] 特許文献1では、発光部と受光部とを有する反射式センサを備えた音声認識装置が開示されている。

[0004] 特許文献2では、音声入力部の角度を検出する角度検出部と、音声入力部とユーザとの距離を検出する距離検出部とを備えた音声入力装置が記載されている。

先行技術文献

特許文献

[0005] 特許文献1：特開平7－64595号公報

特許文献2：特開2010－217754号公報

発明の概要

発明が解決しようとする課題

[0006] 本開示の一態様は、安価な構成で音声認識の誤動作を防止することができる音声認識方法及び音声認識装置を提供する。

課題を解決するための手段

[0007] 本開示の一局面に係る音声認識方法は、機器を動作させるための音声を認識する音声認識方法であって、所定の空間に配置されたマイクロフォンから音声情報を取得する音声情報取得ステップと、前記音声情報取得ステップにおいて取得された前記音声情報に基づいて、前記空間内における音圧の分布を示す空間音圧分布を検出する空間音圧分布検出ステップと、前記空間音圧分布検出ステップにおいて検出された前記空間音圧分布に基づいて、前記空

間内における点音源を検出する点音源検出ステップと、前記点音源検出ステップにおいて前記点音源が検出された場合、前記音声情報取得ステップにおいて取得された前記音声情報に対して音声認識処理を行うと判断する音声認識制御ステップとを含む。

[0008] なお、これらの包括的または具体的な態様は、システム、方法、集積回路、コンピュータプログラム、または、記録媒体で実現されてもよく、システム、装置、方法、集積回路、コンピュータプログラムおよび記録媒体の任意な組み合わせで実現されてもよい。

発明の効果

[0009] 本開示によれば、安価な構成で音声認識の誤動作を防止することができる。

図面の簡単な説明

[0010] [図1]本開示の実施の形態1における音声認識システムの構成を示す図である。

[図2]本開示の実施の形態1における音声認識システムの機器の構成を示す図である。

[図3]本開示の実施の形態1における音声認識部の構成を示す図である。

[図4]本開示の実施の形態1における音声認識システムの動作を説明するためのフローチャートである。

[図5]本開示の実施の形態2における音声認識部の構成を示す図である。

[図6]本開示の実施の形態2における音声認識システムの動作を説明するためのフローチャートである。

[図7]本開示の実施の形態3における音声認識部の構成を示す図である。

[図8]本開示の実施の形態4における音声認識部の構成を示す図である。

[図9]本開示の実施の形態5における音声認識部の構成を示す図である。

[図10]本開示の実施の形態5における音声認識システムの動作を説明するためのフローチャートである。

[図11]本開示の実施の形態6における音声認識部の構成を示す図である。

[図12]本開示の実施の形態7における音声認識部の構成を示す図である。

[図13]本開示の実施の形態8における音声認識部の構成を示す図である。

[図14]本開示の実施の形態8における音声認識システムの動作を説明するためのフローチャートである。

[図15]本開示の実施の形態9における音声認識部の構成を示す図である。

[図16]本開示の実施の形態10における音声認識部の構成を示す図である。

[図17]本開示の実施の形態11における音声認識部の構成を示す図である。

[図18]本開示の実施の形態11において、表示部に表示される表示画面の一例を示す図である。

[図19]本開示の実施の形態12における音声認識部の構成を示す図である。

[図20](A)は、本開示の実施の形態12において、音声認識処理が行われる場合に表示部に表示される表示画面の一例を示す図であり、(B)は、本開示の実施の形態12において、音声認識処理が行われない場合に表示部に表示される表示画面の一例を示す図である。

[図21]本開示の実施の形態13における音声認識部の構成を示す図である。

[図22]本開示の実施の形態13における音声認識システムの動作を説明するためのシーケンス図である。

[図23]本開示の実施の形態13において、表示部に表示される表示画面の一例を示す図である。

[図24](A)は、本開示の実施の形態1において、機器1と複数の音源が存在する空間を模式的に表した図であり、(B)は、本開示の実施の形態1において、空間音圧分布検出部202が検出した空間音圧分布を示した図である。

[図25](A)は、本開示の実施の形態2において、機器1と複数の音源が存在する空間を模式的に表した図であり、(B)は、本開示の実施の形態2において、空間音圧分布検出部202が検出した空間音圧分布を示した図である。

[図26]本開示の実施の形態5において、時刻tにおいて測定された空間内の

音源と、時刻 $t + 1$ において測定された空間内の音源を模式的に示した図である。

発明を実施するための形態

[0011] (本開示の基礎となった知見)

従来、音声によって端末を制御する機能又は音声によってキーワード検索する機能が検討されている。これらの機能を実現する上で、従来、端末を操作するためのリモートコントローラに搭載されたマイクロフォンにより音声を収音していた。近年では、さらなる利便性の向上を目的として、端末がマイクロフォンを内蔵し、端末から離れた場所においてユーザが発話しても端末を動作させることができる音響処理技術が検討されている。

[0012] しかしながら、ユーザが端末から離れた場所において発話する場合、ユーザとマイクロフォンが離れるので、ユーザの発話と周囲の不要な音との音量差がなくなるため、音声認識装置が正しく発話内容を認識することが困難になる。そこで、音声認識装置は、ユーザが端末に対して発した音声か、そうでない音声（例えば、離れたところでの雑談、独り言、又は周辺に存在する家電などが発する音）か、を区別する必要がある。このような課題を解決する手段として、音声入力状態のオン／オフを切り替える手段が検討されている。

[0013] 例えば、特許文献 1 では、音声認識装置は、発光部と受光部とを有する反射式センサを備え、人がいる場合に受光した反射光のレベルと人がいない場合に受光した反射光のレベルとの違いを利用して、ユーザの位置を検出し、音声認識の誤動作を防止している。

[0014] 一方、特許文献 2 では、音声入力装置は、音声入力部の角度を検出する角度検出部と、音声入力部とユーザとの距離を検出する距離検出部とを備える。角度検出部としては、加速度センサ、又は機械式の傾斜センサ等の角度検出器が用いられる。距離検出部としては、位置検出素子、電荷結合素子、又は相補型金属酸化膜半導体等の受光素子を利用した光学センサ、又は超音波センサが用いられる。この従来の音声入力装置では、検出された角度と、検

出された距離とに基づいて、音声入力部のオン／オフが制御され、音声認識の誤動作を防止している。

[0015] すなわち、従来、音声認識装置を構成する上で、誤動作を防止するためには、別途センサを設ける必要があり、これらのセンサを設けることはコストアップに繋がるという課題がある。

[0016] 以上の考察により、本発明者らは本開示の各態様を想到するに至った。

[0017] 以下添付図面を参照しながら、本開示の実施の形態について説明する。なお、以下の実施の形態は、本開示を具体化した一例であって、本開示の技術的範囲を限定するものではない。

[0018] (実施の形態 1)

図 1 は、本開示の実施の形態 1 における音声認識システムの構成を示す図である。図 1 に示す音声認識システムは、機器 1 及びサーバ 2 を備える。

[0019] 機器 1 は、例えば家庭内に配置される家電機器を含む。機器 1 は、ネットワーク 3 を介してサーバ 2 と通信可能に接続される。ネットワーク 3 は、例えばインターネットである。

[0020] なお、機器 1 には、ネットワーク 3 と接続可能な機器（例えば、スマートフォン、パーソナルコンピュータ又はテレビ等）もあれば、それ自身ではネットワーク 3 と接続不可能な機器（例えば、照明機器、洗濯機又は冷蔵庫等）も存在する。それ自身ではネットワーク 3 と接続不可能な機器であっても、ホームゲートウェイを介してネットワーク 3 と接続可能となる機器が存在してもよい。ネットワーク 3 と接続可能な機器は、ホームゲートウェイを介さず、直接サーバ 2 と接続してもよい。

[0021] サーバ 2 は、公知のサーバコンピュータ等から構成され、ネットワーク 3 を介して機器 1 と通信可能に接続されている。

[0022] 図 2 は、本開示の実施の形態 1 における音声認識システムの機器の構成を示す図である。本実施の形態 1 における機器 1 は、通信部 11、制御部 12、メモリ 13、マイクロフォン 14、スピーカ 15 及び表示部 16 を備える。なお、機器 1 のこれらの構成のうち一部の構成がかけていてもよいし、他

の構成を備えていてもよい。

[0023] 通信部 11 は、ネットワーク 3 を介してサーバ 2 へ情報を送信するとともに、ネットワーク 3 を介してサーバ 2 から情報を受信する。制御部 12 は、例えば CPU（中央演算処理装置）で構成され、機器 1 の全体を制御する。制御部 12 は、音声認識部 100 を備える。音声認識部 100 は、ユーザの音声を認識する。制御部 12 は、認識した音声に応じて機器 1 を動作させる。

[0024] メモリ 13 は、例えば ROM（リードオンリメモリ）又は RAM（ランダムアクセスメモリ）で構成され、情報を記憶する。マイクロフォン 14 は、音声を電気信号に変換し、音声情報として出力する。マイクロフォン 14 は、2つのマイクユニット含むマイクロフォンアレイで構成され、機器 1 が配置されている空間内の音声を收音する。スピーカ 15 は、音声を出力する。表示部 16 は、例えば液晶表示装置で構成され、種々の情報を表示する。

[0025] 図 3 は、本開示の実施の形態 1 における音声認識部の構成を示す図である。本実施の形態 1 における音声認識部 100 は、音声取得部 201、空間音圧分布検出部 202、点音源検出部 203、音声認識制御部 204 及び音声認識処理部 205 を備える。

[0026] 音声取得部 201 は、所定の空間に配置されたマイクロフォン 14 から音声情報を取得する。具体的には、マイクロフォン 14 は、アナログ信号である音声をデジタル信号である音声情報に変換し、音声取得部 201 は、デジタル信号に変換された音声情報をマイクロフォン 14 から取得する。なお、所定の空間は、例えば機器が設置されている部屋であることが好ましい。

[0027] 空間音圧分布検出部 202 は、音声取得部 201 によって取得された音声情報に基づいて、空間内における音圧の分布を示す空間音圧分布を検出する。ここで空間音圧分布とは例えば、機器 1（マイクロフォン 14）の位置を基準とした際の、所定角度ごとの音圧の分布を示すものである。また空間音圧分布とは例えば、機器 1（マイクロフォン 14）の位置を基準とした際の、所定角度および所定距離ごとの音圧の分布を示すものであってもよい。空

間音圧分布の具体例については本実施の形態および実施の形態2にて後述する。

[0028] 点音源検出部203は、空間音圧分布検出部202によって検出された空間音圧分布に基づいて、空間内における点音源を検出する。具体的には、点音源検出部203は、まず空間音圧分布検出部202によって検出された空間音圧分布のうち、音圧が所定の閾値以上である音源の角度範囲を取得する。そして、取得した角度範囲の幅が所定の幅以下である音源を、空間内における点音源と判断する。発話する人間の口は点音源であると見なすことができ、騒音を発するエアコンなどの機器は人間の口よりも音源に幅があったり、拡散性もあり、面音源であるから見なすことができる。そのため、空間内における点音源を検出することで、人間が発した音声であるか否かを判断することができる。

[0029] 音声認識制御部204は、点音源検出部203によって点音源が検出された場合、音声取得部201によって取得された音声情報に対して音声認識処理を行うと判断する。

[0030] 音声認識処理部205は、音声認識制御部204によって音声認識処理を行うと判断された場合、音声取得部201によって取得された音声情報に対して音声認識処理を行う。

[0031] ここで、本開示の実施の形態1における音声認識システムの動作について説明する。図4は、本開示の実施の形態1における音声認識システムの動作を説明するためのフローチャートである。

[0032] まず、ステップS1において、音声取得部201は、機器1に設けられたマイクロフォン14から音声情報を取得する。本実施の形態1において、マイクロフォン14は、例えば2つのマイクユニットを含み、2つのマイクユニットから取得された音声情報を音声取得部201に出力する。なおマイクロフォン14に含まれるマイクユニットは2つに限らず、3つ以上含まれていてもよい。

[0033] 次に、ステップS2において、空間音圧分布検出部202は、音声取得部

201によって取得された音声情報に基づいて、空間内における音圧の分布を示す空間音圧分布を検出する。

[0034] 例えば、マイクロフォン14が複数のマイクユニットを含む場合、既知のビームフォーミング技術などによって收音範囲に指向性を持たせることができるので、指向性のビーム方向を振る（走査させる）ことによって所定角度ごとの音圧分布を検出できる。さらに、音波が球面波であるという音波の特徴を利用することで、音源の距離に関するビーム操作も可能となり、所定の距離ごとの音圧分布の検出も可能となる。

[0035] 図24（A）は機器1と複数の音源（音源241、音源242、音源243、音源244）が存在する空間を模式的に示した図であり、図24（B）は図24（A）に示した空間において空間音圧分布検出部202が検出した空間音圧分布を示した図である。

[0036] 図24（A）に示す機器1のマイクロフォン14は、水平方向に位置の異なるマイクユニット14aとマイクユニット14bとを含む。また図24（A）に示すように空間内に複数の音源（音源241、音源242、音源243、音源244）が存在するものとする。複数のマイクユニットを含むマイクロフォン14は上記のように收音範囲に指向性を持たせることができるので、指向性のビーム方向を振る（走査させる）ことによって空間内の複数の領域に対してそれぞれ音源の存在の有無を検出することが出来る。これによって、図24（B）に示すような空間音圧分布を検出することが出来る。図24（B）に示す空間音圧分布は、機器1の位置を0°とした際に、-90°～90°における角度ごとの音圧の分布を示すものである。

[0037] なお、マイクロフォン14が水平方向に位置の異なる2つのマイクユニットを含む場合は、図24に示すように水平面上での空間音圧分布を検出することが出来る。また、マイクロフォン14が垂直方向に位置の異なる2つのマイクユニットを含む場合は、垂直面上での空間音圧分布を検出することが出来る。また、マイクロフォン14が水平・垂直方向に位置が異なる3つのマイクユニットを含む場合、マイクロフォン14の位置を原点とする3次元

座標空間における音圧レベルの分布を検出することが出来る。

[0038] 次に、ステップS3において、点音源検出部203は、空間音圧分布検出部202によって検出された空間音圧分布に基づいて、空間内における点音源を検出する。点音源の幅は面音源の幅よりも狭い。また、点音源の面積は、面音源の面積よりも小さい。そのため、点音源検出部203は、まず空間音圧分布検出部202によって検出された空間音圧分布のうち、音圧が所定の閾値以上である音源の角度範囲を取得する。そして、取得した角度範囲の幅が所定の幅以下である音源を、空間内における点音源と判断する。以下図24を用いて点音源検出部203が空間内における点音源を特定する具体例を説明する。

[0039] 点音源検出部203は、まず図24(B)に示す空間音圧分布のうち、音圧が所定の閾値245以上である角度範囲を特定する。これにより点音源か否かを判断する対象となる音源を特定する。図24(B)の下部には $-90^{\circ} \sim 45^{\circ}$ の範囲内に音源241が、 $-45^{\circ} \sim 0^{\circ}$ の範囲内に音源242が、 $0^{\circ} \sim 45^{\circ}$ の範囲内に音源243が、 $45^{\circ} \sim 90^{\circ}$ の範囲内に音源244が、それぞれ点音源か否かを判断する対象となる音源として特定されている。また空間音圧分布検出部202は、音源241の角度範囲A、音源242の角度範囲B、音源243の角度範囲C、音源244の角度範囲Dをそれぞれ取得している。

[0040] そして点音源検出部203は、対象となる音源として特定した音源241、音源242、音源243、音源244のそれぞれの角度範囲A～Dの幅のうち、所定の幅以下である角度範囲を持つ音源を特定し、特定した音源を点音源として判断する。ここでは例えば所定の幅以下である角度範囲BおよびDに対応する音源242・音源244を、点音源であると判断する。これによって図24(A)に示す空間内には点音源242・点音源244が存在することを判断することができる。

[0041] なお、点音源検出部203が点音源であるか否かを判断する方法は、上で説明した方法に限られない。点音源検出部203は取得した音源に関する、

音圧の大小や、音圧の幅（すなわち特定した音源の角度範囲）、音波の特徴、音圧の面積（実施の形態２で説明）などのあらゆる情報のうち一または複数の情報を取得することで点音源であるか否かを判断する。

[0042] 次に、ステップＳ４において、音声認識制御部２０４は、点音源検出部２０３によって点音源が検出されたか否かを判断する。ここで、点音源が検出されたと判断された場合（ステップＳ４でＹＥＳ）、音声認識制御部２０４は、音声認識処理を行うと判断し、ステップＳ５の処理へ移行する。一方、点音源が検出されていないと判断された場合（ステップＳ４でＮＯ）、音声認識制御部２０４は、音声認識処理を行わないと判断し、ステップＳ１の処理へ戻る。図２４の例では、点音源検出部２０３は点音源２４２・点音源２４４を検出しているので、ステップＳ５の処理へ移行する。

[0043] なお図２４（Ｂ）のように、所定角度ごとの音圧分布を検出している場合は、マイクロフォン１４から所定の角度範囲内に位置する音源のみを制御対象の音源と判断してもよい。例えば音声認識制御部２０４は、音源２４２・音源２４４のうち、マイクロフォン１４から所定の角度範囲（例えば−４５°～４５°）内に位置する音源である音源２４２のみを制御対象の音源と判断する。これによって、音声認識装置（システム）が音声コマンドを受け付ける音源領域を制限することが出来る。なお、音声認識制御部２０４が点音源と判断する所定の角度範囲に関しては、予め設定されている角度でもよいし、ユーザが設定出来てもよい。

[0044] 次に、ステップＳ５において、音声認識処理部２０５は、音声認識制御部２０４によって音声認識処理を行うと判断された場合、音声取得部２０１によって取得された音声情報に対して音声認識処理を行う。また、音声認識処理部２０５は、音声認識制御部２０４によって音声認識処理を行わないと判断された場合、音声取得部２０１によって取得された音声情報に対して音声認識処理を行わない。

[0045] このように、点音源が検出された場合、音声認識処理が行われ、点音源が検出されない場合、音声認識処理が行われないので、音声認識処理を行うか

否かを容易に判断することができる。また、人間の存在を検知するためのセンサが不要であるので、安価な構成で音声認識の誤動作を防止することができる。

[0046] なお、本実施の形態において、マイクロフォン１４は、２つのマイクユニットを含んでいるが、本開示は特にこれに限定されず、マイクロフォン１４は、３つ以上のマイクユニットを含んでもよい。この場合、空間音圧分布の検出精度を向上させることができる。このことは、他の実施の形態でも適用可能である。

[0047] また、本実施の形態において、機器１が備えるマイクロフォン１４を利用して空間音圧分布が検出されるが、本開示は特にこれに限定されず、空間音圧分布検出部２０２は、機器１が備えるマイクロフォン１４に加えて、機器１と同一空間内に配置された他の機器が備えるマイクロフォンを利用して空間音圧分布を検出してもよい。この場合、他の機器のマイクロフォンの位置は、空間音圧分布検出部２０２に予め記憶されていることが好ましい。このことは、他の実施の形態でも適用可能である。

[0048] また、本実施の形態において、機器１が音声認識部１００を備えているが、本開示は特にこれに限定されず、サーバ２が音声認識部１００を備えてもよい。この場合、機器１のマイクロフォン１４で取得された音声情報が、ネットワーク３を介してサーバ２へ送信され、サーバ２の音声認識部１００が図４のステップＳ１～Ｓ５の処理を実行する。そして、サーバ２は、音声認識処理を行った場合、音声認識処理結果に基づく機器１の動作指示を機器１へ送信し、機器１は、サーバ２からの動作指示に応じて動作する。このことは、他の実施の形態でも適用可能である。

[0049] また、本実施の形態において、機器１が音声認識部１００を備えているが、本開示は特にこれに限定されず、機器１と同一の家庭内に配置された他の機器が音声認識部１００を備えてもよい。また、家庭内に配置された複数の機器を制御するホームコントローラが音声認識部１００を備えてもよい。このことは、他の実施の形態でも適用可能である。

[0050] (実施の形態2)

図5は、本開示の実施の形態2における音声認識部の構成を示す図である。本実施の形態2における音声認識部101は、音声取得部201、空間音圧分布検出部202、点音源検出部203、音声認識制御部204、音声認識処理部205及び点音源位置特定部206を備える。なお、機器1の構成は、実施の形態1と同じであり、制御部12が音声認識部101を備える。実施の形態2において、実施の形態1と同一の構成については同一の符号を付し、説明を省略する。

[0051] 点音源位置特定部206は、点音源検出部203によって点音源が検出された場合、検出された点音源の空間内における位置を特定する。なお、メモリ13は、機器1に設けられたマイクロフォンに含まれる複数のマイクユニットの位置を予め記憶している。点音源位置特定部206は、複数のマイクユニットの位置を用いた3点測量法により3次元空間内における点音源の位置を特定する。

[0052] 音声認識制御部204は、点音源の位置と機器1との距離が所定の距離より近い場合、音声認識処理を行うと判断し、点音源の位置と機器1との距離が所定の距離以上である場合、音声認識処理を行わないと判断する。

[0053] ここで、本開示の実施の形態2における音声認識システムの動作について説明する。図6は、本開示の実施の形態2における音声認識システムの動作を説明するためのフローチャートである。

[0054] なお、図6のステップS11～S14の処理は、図4のステップS1～S4の処理と同じであるので説明を省略する。

[0055] ステップS14において点音源が検出されたと判断された場合（ステップS4でYES）、ステップS15において、点音源位置特定部206は、点音源検出部203によって検出された点音源の空間内における位置を特定する。点音源位置特定部206は、空間音圧分布検出部202によって検出された空間音圧分布の中から、点音源検出部203によって検出された点音源の位置を特定する。

[0056] 空間音圧分布は、例えば、マイクロフォン14の位置を原点とする2次元（もしくは3次元）座標空間における音圧レベルの分布を表している。メモリ13は、機器1に設けられたマイクロフォン14に含まれる複数のマイクユニットの位置を予め記憶している。また、複数のマイクユニットは、所定の間隔で配置されている。点音源から出力された音声は各マイクユニットに到達するまでの時間は、それぞれ異なる。そのため、点音源位置特定部206は、複数のマイクユニットのそれぞれの位置と、検出された点音源から複数のマイクユニットに音声は到達するまでの時間とに基づいて、2次元（3次元）空間内における点音源の位置（マイクロフォン14からの距離）を特定する。

[0057] ここで、点音源の位置（マイクロフォン14からの距離）を特定する方法は、これに限られない。例えば音波が伝わる特徴を利用して特定してもよい。例えば音の伝播が球面波であるという特徴を利用することで距離方向の音圧分布を測定することが出来るので、これにより点音源の位置（マイクロフォン14からの距離）を特定してもよい。

[0058] ここで、ステップS15にて、点音源の位置を特定した場合、空間音圧分布検出部は、ステップS12にて検出した音圧分布を機器1を基準にした所定角度および所定距離ごとの音圧分布を示すものに更新してもよい。

[0059] 図25（A）は機器1と複数の音源（音源251、音源252、音源253、音源254）が存在する空間を模式的に示した図であり、図25（B）は図25（A）に示した空間において空間音圧分布検出部202が検出した距離と角度ごとの空間音圧分布を示した図である。図25（B）に示す空間音圧分布は、機器1の位置を原点とした際に、横軸：角度、縦軸：距離として2次元座標を構築し、検出した音源を座標上に配置し、また各音源の面積（角度範囲の幅および奥行き方向の幅）を円の大きさにて表したものである。

[0060] ステップS15にて音源の位置（距離）を特定しているので、各音源に対して機器1から見た奥行き方向の幅を検出することが出来る。よって、ステ

ップS 13にて、実施の形態1で説明した各音源の角度範囲の幅に加えて、奥行き方向の幅を考慮することで、各音源を面積として算出できる。これによりこの面積が所定の閾値よりも小さいものを点音源として特定してもよい。

[0061] 図25(B)に示す例では、 $-90^{\circ} \sim 45^{\circ}$ の範囲内に音源251が、 $-45^{\circ} \sim 0^{\circ}$ の範囲内に音源252が、 $0^{\circ} \sim 45^{\circ}$ の範囲内に音源253が、 $45^{\circ} \sim 90^{\circ}$ の範囲内に音源254が、それぞれ音源として検出されており、これらの音源のうち所定の面積よりも小さい音源である音源252、音源253、音源254が点音源と特定されているものとする。

[0062] 次に、ステップS 16において、音声認識制御部204は、点音源の位置と機器1との距離が所定の距離より近いかな否かを判断する。音声認識制御部204は、マイクロフォン14の位置を原点とする3次元座標空間における機器1の位置を予め記憶している。そのため、音声認識制御部204は、点音源位置特定部206によって特定された点音源の位置座標と、機器1の位置座標との間の距離を算出することができ、算出した距離が所定の距離より近いかな否かを判断する。

[0063] ここで図25(B)を用いて、音声認識制御部204が点音源の位置と機器1との距離が所定の距離より近いかな否かを判断する例について説明する。

[0064] 図25(B)の例では上記したように点音源検出部203が音源252、音源253、音源254、を点音源であると判断したとする。この場合、点音源252・点音源253・点音源254のうち、所定の距離である閾値255(この場合3m)内に位置する点音源252・点音源254が、機器1との距離が所定の範囲内に納まる音源であると判断する。

[0065] なお、所定の距離とは、例えば、ユーザが機器1を操作可能な範囲であることが好ましい。例えば、機器1がテレビである場合、所定の距離は、ユーザが視聴可能な範囲であることが好ましい。また、マイクユニットの検出感度を考慮した場合、所定の距離とは、複数のマイクユニットのうち的一方端のマイクユニットから他方端のマイクユニットまでの距離の3倍、4倍又は

5倍であることが好ましい。

[0066] ここで、点音源の位置と機器1との距離が所定の距離より近いと判断された場合（ステップS16でYES）、音声認識制御部204は、音声認識処理を行うと判断し、ステップS17の処理へ移行する。一方、点音源の位置と機器1との距離が所定の距離より近くないと判断された場合（ステップS16でNO）、音声認識制御部204は、音声認識処理を行わないと判断し、ステップS11の処理へ戻る。

[0067] 次に、ステップS17において、音声認識処理部205は、音声認識制御部204によって音声認識処理を行うと判断された場合、音声取得部201によって取得された音声情報に対して音声認識処理を行う。また、音声認識処理部205は、音声認識制御部204によって音声認識処理を行わないと判断された場合、音声取得部201によって取得された音声情報に対して音声認識処理を行わない。

[0068] このように、点音源が検出された場合、点音源の位置と機器1との距離が所定の距離より近いかな否かが判断される。そして、点音源の位置と機器1との距離が所定の距離より近いと判断された場合、音声認識処理が行われ、点音源の位置と機器1との距離が所定の距離より近くないと判断された場合、音声認識処理が行われな。したがって、点音源の位置と機器1との距離が所定の距離より近いかな否かを判断することにより、機器1を操作するためにユーザが機器1に近い場所で発話しているのか、又は機器1の操作とは関係なくユーザが機器1から離れた場所で発話しているのかを判断することができ、安価な構成で音声認識の誤動作を防止することができる。

[0069] また、高価な装置を用いることなく話者の位置を特定し、話者の位置に基づいて音声認識処理を制御することができる。

[0070] （実施の形態3）

図7は、本開示の実施の形態3における音声認識部の構成を示す図である。本実施の形態3における音声認識部102は、音声取得部201、空間音圧分布検出部202、点音源検出部203、音声認識制御部204、音声認

識処理部 205、点音源位置特定部 206 及び指向性制御部 207 を備える。なお、機器 1 の構成は、実施の形態 1 と同じであり、制御部 12 が音声認識部 102 を備える。実施の形態 3 において、実施の形態 1 及び実施の形態 2 と同一の構成については同一の符号を付し、説明を省略する。

[0071] 指向性制御部 207 は、点音源位置特定部 206 によって特定された点音源の位置に向けてマイクロフォン 14 の指向性の方向を制御する。指向性制御部 207 は、点音源位置特定部 206 によって特定された点音源の位置に向かう方向の感度を高めるように、音声取得部 201 によって取得された音声情報に対して信号処理を施す。これにより、マイクロフォン 14 の指向性が、点音源に向かう方向に制御される。

[0072] このように、点音源の位置に向けてマイクロフォン 14 の指向性の方向を制御することにより、点音源からの音声情報を確実に取得することができ、音声認識の誤動作を防止することができる。

[0073] (実施の形態 4)

図 8 は、本開示の実施の形態 4 における音声認識部の構成を示す図である。本実施の形態 4 における音声認識部 103 は、音声取得部 201、空間音圧分布検出部 202、点音源検出部 203、音声認識制御部 204、音声認識処理部 205、点音源位置特定部 206、指向性制御部 207、出力音声生成部 208 及びエコー除去部 209 を備える。なお、機器 1 の構成は、実施の形態 1 と同じであり、制御部 12 が音声認識部 103 を備える。実施の形態 4 において、実施の形態 1 ～実施の形態 3 と同一の構成については同一の符号を付し、説明を省略する。

[0074] 出力音声生成部 208 は、機器 1 のスピーカ 15 から出力される出力音声情報を生成する。出力音声生成部 208 は、ユーザに対して操作を促すための出力音声情報、又はユーザの操作に対して応答するための出力音声情報などを生成し、スピーカ 15 及びエコー除去部 209 へ出力する。出力音声生成部 208 は、機器 1 の動作状況に応じた出力音声情報をメモリ 13 から読み出し、出力音声情報を生成する。あるいは、出力音声生成部 208 は、機

器 1 の動作状況に応じた複数の音声情報をメモリ 13 から読み出し、読み出した複数の音声情報を組み合わせることにより出力音声情報を生成してもよい。なお、例えば機器 1 がテレビ、ラジオ又は音楽再生用のオーディオなどである場合、出力音声生成部 208 が生成する出力音声情報としては、受信した放送コンテンツの音声情報又は記録（録画又は録音）されたコンテンツの音声情報を含んでいてもよい。

[0075] スピーカ 15 は、出力音声生成部 208 によって生成された出力音声情報を出力音声として出力する。

[0076] エコー除去部 209 は、出力音声生成部 208 によって生成された出力音声情報を取得し、音声取得部 201 によって取得された音声情報から、取得した出力音声情報をエコー成分として除去する。

[0077] すなわち、スピーカ 15 から音声が出力される場合、音声取得部 201 によって取得された音声情報には、ユーザが発した音声だけでなく、当該スピーカ 15 から出力された音声のエコー成分として含まれる。そのため、エコー成分が含まれる音声情報を基に音声認識が行われた場合、正確に音声認識が行われないおそれがある。そこで、音声取得部 201 によって取得された音声情報から、出力音声生成部 208 によって生成された出力音声情報をエコー成分として除去する。これにより、点音源からの音声情報を確実に取得することができ、音声認識の誤動作を防止することができる。

[0078] （実施の形態 5）

図 9 は、本開示の実施の形態 5 における音声認識部の構成を示す図である。本実施の形態 5 における音声認識部 104 は、音声取得部 201、空間音圧分布検出部 202、点音源検出部 203、音声認識制御部 204、音声認識処理部 205、点音源位置特定部 206 及び直接到来点音源特定部 210 を備える。なお、機器 1 の構成は、実施の形態 1 と同じであり、制御部 12 が音声認識部 104 を備える。実施の形態 5 において、実施の形態 1～実施の形態 4 と同一の構成については同一の符号を付し、説明を省略する。

[0079] 直接到来点音源特定部 210 は、点音源位置特定部 206 によって位置が

特定された点音源のうち、発せられた音声マイクロフォン14に直接到来する直接到来点音源を特定する。なお、点音源位置特定部206によって位置が特定された点音源のうち、発せられた音声マイクロフォン14に直接到来する直接到来点音源は、周知の方法で特定することが可能である。直接到来音源の特定方法に関しては図26を用いて後述する。音声認識制御部204は、直接到来点音源特定部210で直接到来点音源が特定された場合、音声認識処理を行うと判断し、直接到来点音源特定部210で直接到来点音源が特定されない場合、音声認識処理を行わないと判断する。

[0080] ここで、本開示の実施の形態5における音声認識システムの動作について説明する。図10は、本開示の実施の形態5における音声認識システムの動作を説明するためのフローチャートである。

[0081] なお、図10のステップS21～S25の処理は、図6のステップS11～S15の処理と同じであるので説明を省略する。

[0082] ステップS26において、直接到来点音源特定部210は、点音源位置特定部206によって位置が特定された点音源のうち、発せられた音声マイクロフォン14に直接到来する直接到来点音源を特定する。

[0083] 次に、ステップS27において、音声認識制御部204は、直接到来点音源特定部210によって直接到来点音源が特定されたか否かを判断する。

[0084] 点音源からマイクロフォン14に到来する音声には、直接マイクロフォン14に到来する音声だけでなく、壁などに反射してからマイクロフォン14に到来する音声又は壁などを通過してからマイクロフォン14に到来する音声もある。

[0085] 通常、ユーザは、機器の方向を向いた状態で機器を操作する。そのため、壁などに反射してからマイクロフォン14に到来する音声又は壁などを通過してからマイクロフォン14に到来する音声は、ユーザが機器の方向を向いていない状態で発せられた音声である可能性が高く、機器に対する操作指示ではない可能性が高い。そこで、本実施の形態5における音声認識処理部205は、発せられた音声マイクロフォン14に直接到来する場合、音声認

識を行い、発せられた音声マイクロフォン 14 に直接到来しない場合、音声認識を行わない。

[0086] 以下図 26 を用いて、直接到来音源特定部 210 が直接到来音か直接到来音でない音（反射音）かを判断する方法を説明する。図 26 は、時刻 t において測定された空間内の音源と、時刻 $t + 1$ において測定された空間内の音源を模式的に示した図である。

[0087] 図 26 の例では、点音源検出部 203 は、時刻 t において点音源 261 を検出している。また時刻 t において、点音源位置特定部 206 は点音源 261 の位置が、機器 1 から距離 x の位置であることを特定している。また時刻 $t + 1$ において、点音源検出部 203 は、点音源 262 を、点音源 263 をそれぞれ検出している。また時刻 $t + 1$ において、点音源位置特定部 206 は点音源 262 の位置が機器 1 から距離 y の位置であり、点音源 263 の位置が機器 1 から距離 x の位置であることを特定している。また、点音源位置特定部 206 は点音源 261 と点音源 263 が同じ位置（2次元座標）にあることを特定している。

[0088] この場合、まず直接到来音源特定部 210 は、異なる時刻に同じ位置（2次元座標）に発生している点音源 261 と点音源 263 を同じ音源から発せられた音と判断する。

[0089] そして次に直接到来音源特定部 210 は、異なる時刻に検出された音源のうち、同じ音源から発せられた音と判断されていない点音源 261 と点音源 262 との音の特徴を比較する。ここで、音の特徴とは例えば、時刻 t において点音源 261 に指向性を向けた際の音声と時刻 $t + 1$ において点音源 262、もしくは点音源 263 に指向性を向けた際の音声との相関や独立性などを含み、特に限定しない。そして直接到来音源特定部 210 は、点音源 261 と点音源 262 が略同一の音の特徴を有していると判断した場合、点音源 262 を直接到来音源でない、すなわち点音源 261 が壁に反射することによって発生している反射音源であると判断する。そして音源 261、音源 263 に関しても、機器 1 に向かって発されている音源ではない可能性が高

いので、直接到来音源ではないと判断する。

[0090] すなわち直接到来音源特定部 210 は、異なる時刻に検出され検出位置（音源分布上の検出領域）が異なる複数の音源のうち、略同一の音の特徴を有する複数の音源を検出した際に、少なくとも検出した時刻が後である音源を直接到来音でない音と判断する。

[0091] ここで音の特徴が略同一であるか否かの判断の方法としては、例えば複数の音源の音の特徴を示す値を比較し、それぞれの値の一致度が所定値以上である場合に略同一として判断するなどが考えられる。

[0092] なお、直接到来音源の特定方法はこれに限られない。例えば、同一時刻に音の特徴が略同一である複数の音源を特定した際に、機器からの距離が遠い音源を直接到来音ではない音と判断してもよい。

[0093] すなわち、直接到来音源特定部 210 は時間的に異なる（もしくは時間的に同一の）複数の音源が存在する場合に、各々の音源の機器 1 に対する距離や、位置、音の特徴、などに基づいて直接到来音であるか否かを判断すればよい。

[0094] ここで、直接到来点音源特定部 210 によって直接到来点音源が特定されたと判断された場合（ステップ S27 で YES）、音声認識制御部 204 は、音声認識処理を行うと判断し、ステップ S28 の処理へ移行する。一方、直接到来点音源特定部 210 によって直接到来点音源が特定されないと判断された場合（ステップ S27 で NO）、音声認識制御部 204 は、音声認識処理を行わないと判断し、ステップ S21 の処理へ戻る。

[0095] 次に、ステップ S28 において、音声認識処理部 205 は、音声認識制御部 204 によって音声認識処理を行うと判断された場合、音声取得部 201 によって取得された音声情報に対して音声認識処理を行う。また、音声認識処理部 205 は、音声認識制御部 204 によって音声認識処理を行わないと判断された場合、音声取得部 201 によって取得された音声情報に対して音声認識処理を行わない。

[0096] このように、発せられた音声がマイクロフォン 14 に直接到来する場合、

音声認識が行われ、発せられた音声マイクロフォン 14 に直接到来しない場合、音声認識が行われないので、機器 1 を操作するためにユーザが機器 1 の方向に向かって発話しているのか、又は機器 1 の操作とは関係なくユーザが機器 1 の方向とは異なる方向に向かって発話しているのかを判断することができ、安価な構成で音声認識の誤動作を防止することができる。

[0097] (実施の形態 6)

図 11 は、本開示の実施の形態 6 における音声認識部の構成を示す図である。本実施の形態 6 における音声認識部 105 は、音声取得部 201、空間音圧分布検出部 202、点音源検出部 203、音声認識制御部 204、音声認識処理部 205、点音源位置特定部 206、指向性制御部 207 及び直接到来点音源特定部 210 を備える。なお、機器 1 の構成は、実施の形態 1 と同じであり、制御部 12 が音声認識部 105 を備える。実施の形態 6 において、実施の形態 1 ～実施の形態 5 と同一の構成については同一の符号を付し、説明を省略する。

[0098] 直接到来点音源特定部 210 は、点音源位置特定部 206 によって位置が特定された点音源のうち、発せられた音声マイクロフォン 14 に直接到来する直接到来点音源を特定する。

[0099] 指向性制御部 207 は、直接到来点音源特定部 210 によって直接到来点音源が特定された場合、点音源位置特定部 206 によって特定された点音源の位置に向けてマイクロフォン 14 の指向性の方向を制御する。

[0100] このように、発せられた音声マイクロフォン 14 に直接到来する直接到来点音源の位置に向けてマイクロフォン 14 の指向性の方向が制御されるので、点音源からの音声情報を確実に取得することができ、音声認識の誤動作を防止することができる。

[0101] (実施の形態 7)

図 12 は、本開示の実施の形態 7 における音声認識部の構成を示す図である。本実施の形態 7 における音声認識部 106 は、音声取得部 201、空間音圧分布検出部 202、点音源検出部 203、音声認識制御部 204、音声

認識処理部 205、点音源位置特定部 206、指向性制御部 207、出力音声生成部 208、エコー除去部 209 及び直接到来点音源特定部 210 を備える。なお、機器 1 の構成は、実施の形態 1 と同じであり、制御部 12 が音声認識部 106 を備える。実施の形態 7 において、実施の形態 1 ～実施の形態 6 と同一の構成については同一の符号を付し、説明を省略する。

[0102] エコー除去部 209 は、出力音声生成部 208 によって生成された出力音声情報を取得し、音声取得部 201 によって取得された音声情報から、取得した出力音声情報をエコー成分として除去する。

[0103] 直接到来点音源特定部 210 は、点音源位置特定部 206 によって位置が特定された点音源のうち、発せられた音声がマイクロフォン 14 に直接到来する直接到来点音源を特定する。

[0104] 指向性制御部 207 は、直接到来点音源特定部 210 によって直接到来点音源が特定された場合、点音源位置特定部 206 によって特定された点音源の位置に向けてマイクロフォン 14 の指向性の方向を制御する。

[0105] このように、発せられた音声がマイクロフォン 14 に直接到来する直接到来点音源の位置に向けてマイクロフォン 14 の指向性の方向が制御されるので、点音源からの音声情報を確実に取得することができ、音声認識の誤動作を防止することができる。

[0106] また、音声取得部 201 によって取得された音声情報から、出力音声生成部 208 によって生成された出力音声情報がエコー成分として除去されるので、点音源からの音声情報を確実に取得することができ、音声認識の誤動作を防止することができる。

[0107] (実施の形態 8)

図 13 は、本開示の実施の形態 8 における音声認識部の構成を示す図である。本実施の形態 8 における音声認識部 107 は、音声取得部 201、空間音圧分布検出部 202、点音源検出部 203、音声認識制御部 204、音声認識処理部 205、点音源位置特定部 206、直接到来点音源特定部 210 及び音声識別部 211 を備える。なお、機器 1 の構成は、実施の形態 1 と同

じであり、制御部 12 が音声認識部 107 を備える。実施の形態 8 において、実施の形態 1 ～実施の形態 7 と同一の構成については同一の符号を付し、説明を省略する。

[0108] 音声識別部 211 は、直接到来点音源特定部 210 によって直接到来点音源が特定された場合、直接到来した音声人間によって発話された音声であるか否かを識別する。

[0109] 音声認識制御部 204 は、音声識別部 211 によって直接到来した音声人間によって発話された音声であると識別された場合、音声認識処理を行うと判断し、音声識別部 211 によって直接到来した音声人間によって発話された音声ではないと識別された場合、音声認識処理を行わないと判断する。

[0110] 人間の口は点音源であるが、様々な雑音の中には点音源である雑音も存在する。そのため、直接到来した音声人間によって発話された音声であるか否かを識別することにより、人間以外の機器から発せられる雑音に対して音声認識処理が行われることを防止する。

[0111] ここで、本開示の実施の形態 8 における音声認識システムの動作について説明する。図 14 は、本開示の実施の形態 8 における音声認識システムの動作を説明するためのフローチャートである。

[0112] なお、図 14 のステップ S31 ～ S37 の処理は、図 10 のステップ S21 ～ S27 の処理と同じであるので説明を省略する。

[0113] ステップ S38 において、音声識別部 211 は、直接到来点音源特定部 210 によって直接到来点音源が特定された場合、直接到来した音声人間によって発話された音声であるか否かを識別する。メモリ 13 は、人間の音声の波形パターンを予め記憶している。音声識別部 211 は、取得した音声情報の波形パターンと、メモリ 13 に予め記憶されている人間の音声の波形パターンとを比較する。取得した音声情報の波形パターンが、メモリ 13 に予め記憶されている人間の音声の波形パターンと略一致する場合、音声識別部 211 は、直接到来した音声人間によって発話された音声であると識別す

る。一方、取得した音声情報の波形パターンが、メモリ 13 に予め記憶されている人間の音声の波形パターンと略一致しない場合、音声識別部 211 は、直接到来した音声人間によって発話された音声ではないと識別する。

[0114] ここで、直接到来した音声人間によって発話された音声であると識別された場合（ステップ S38 で YES）、音声認識制御部 204 は、音声認識処理を行うと判断し、ステップ S39 の処理へ移行する。一方、直接到来した音声人間によって発話された音声ではないと識別された場合（ステップ S38 で NO）、音声認識制御部 204 は、音声認識処理を行わないと判断し、ステップ S31 の処理へ戻る。

[0115] 次に、ステップ S39 において、音声認識処理部 205 は、音声認識制御部 204 によって音声認識処理を行うと判断された場合、音声取得部 201 によって取得された音声情報に対して音声認識処理を行う。また、音声認識処理部 205 は、音声認識制御部 204 によって音声認識処理を行わないと判断された場合、音声取得部 201 によって取得された音声情報に対して音声認識処理を行わない。

[0116] このように、点音源から直接到来した音声人間の音声である場合、音声認識が行われ、点音源から直接到来した音声人間の音声ではない場合、音声認識が行われないので、人間以外の機器から発せられる音声に対して音声認識処理が行われることを防止することができる。

[0117] なお、音声識別部 211 は、直接到来点音源特定部 210 によって直接到来点音源が特定された場合、話者を識別してもよい。話者を識別する方法については限定しない。音声識別部 211 は、音声を発している話者の大まかな年齢又は性別などの個人情報を特定してもよいし、予め登録されている音声データベース等を用いて個人を特定してもよい。これによって、例えば室外から発生して壁を通過してマイクロフォン 14 に到来した音声、又は音声データベースに登録されている家族以外が発した音声に対して音声認識処理が行われることを防止することができる。

[0118] （実施の形態 9）

図15は、本開示の実施の形態9における音声認識部の構成を示す図である。本実施の形態9における音声認識部108は、音声取得部201、空間音圧分布検出部202、点音源検出部203、音声認識制御部204、音声認識処理部205、点音源位置特定部206、指向性制御部207、直接到来点音源特定部210及び音声識別部211を備える。なお、機器1の構成は、実施の形態1と同じであり、制御部12が音声認識部108を備える。実施の形態9において、実施の形態1～実施の形態8と同一の構成については同一の符号を付し、説明を省略する。

[0119] 音声識別部211は、直接到来点音源特定部210によって直接到来点音源が特定された場合、直接到来した音声人間によって発話された音声であるか否かを識別する。

[0120] 指向性制御部207は、直接到来点音源特定部210によって直接到来点音源が特定されるとともに、直接到来した音声人間によって発話された音声であると識別された場合、点音源位置特定部206によって特定された点音源の位置に向けてマイクロフォン14の指向性の方向を制御する。

[0121] このように、人間であると識別された点音源の位置に向けてマイクロフォン14の指向性の方向が制御されるので、ユーザからの音声情報を確実に取得することができ、音声認識の誤動作を防止することができる。

[0122] (実施の形態10)

図16は、本開示の実施の形態10における音声認識部の構成を示す図である。本実施の形態10における音声認識部109は、音声取得部201、空間音圧分布検出部202、点音源検出部203、音声認識制御部204、音声認識処理部205、点音源位置特定部206、指向性制御部207、出力音声生成部208、エコー除去部209、直接到来点音源特定部210及び音声識別部211を備える。なお、機器1の構成は、実施の形態1と同じであり、制御部12が音声認識部109を備える。実施の形態10において、実施の形態1～実施の形態9と同一の構成については同一の符号を付し、説明を省略する。

[0123] 音声識別部 211 は、直接到来点音源特定部 210 によって直接到来点音源が特定された場合、直接到来した音声人間によって発話された音声であるか否かを識別する。

[0124] 指向性制御部 207 は、直接到来点音源特定部 210 によって直接到来点音源が特定されるとともに、直接到来した音声人間によって発話された音声であると識別された場合、点音源位置特定部 206 によって特定された点音源の位置に向けてマイクロフォン 14 の指向性の方向を制御する。

[0125] 特に、本実施の形態では、指向性制御部 207 は、エコー除去部 209 によってエコー成分が除去された音声情報に対して、点音源の位置に向かう方向に指向性を有するように信号処理を施す。

[0126] これにより、点音源からの音声情報を確実に取得することができ、音声認識の誤動作を防止することができる。

[0127] (実施の形態 11)

図 17 は、本開示の実施の形態 11 における音声認識部の構成を示す図である。本実施の形態 11 における音声認識部 110 は、音声取得部 201、空間音圧分布検出部 202、点音源検出部 203、音声認識制御部 204、音声認識処理部 205、点音源位置特定部 206 及び音声認識用単語表示制御部 212 を備える。なお、機器 1 の構成は、実施の形態 1 と同じであり、制御部 12 が音声認識部 110 を備える。実施の形態 11 において、実施の形態 1 ～ 実施の形態 10 と同一の構成については同一の符号を付し、説明を省略する。

[0128] 音声認識用単語表示制御部 212 は、音声認識制御部 204 によって音声認識処理を行うと判断された場合、音声認識処理に用いられる複数の音声認識用単語を予め記憶するメモリ 13 から、機器の操作に関する少なくとも 1 つの音声認識用単語を読み出し、読み出した少なくとも 1 つの音声認識用単語を表示部 16 に表示する。

[0129] 音声認識用単語は、音声認識が可能な単語である。例えば、機器 1 がテレビである場合、音声認識用単語は、“音量大”、“音量小”、“チャンネル

アップ”、“チャンネルダウン”、“番組表”、“検索”、“動画”及び“電源オフ”を含む。

[0130] 図18は、本開示の実施の形態11において、表示部に表示される表示画面の一例を示す図である。

[0131] 機器1がテレビである場合、表示部16の表示画面300には、図18に示す音声認識用単語301が表示される。図18に示すように、音声認識用単語表示制御部212は、“音量大”、“音量小”、“チャンネルアップ”、“チャンネルダウン”、“番組表”、“検索”、“動画”及び“電源オフ”という音声認識用単語301を表示部16に表示する。なお、音声認識用単語表示制御部212によって表示される音声認識用単語301はこれに限られない。

[0132] このように、音声認識処理に用いられる音声認識用単語が表示されるので、ユーザは表示された音声認識用単語を確認し、確実に音声により機器を操作することができる。

[0133] なお、本実施の形態11では、表示画面300の下部に音声認識用単語301が一覧表示されているが、本開示は特にこれに限定されず、表示画面300の上部、中央部、左端部又は右端部に音声認識用単語301が一覧表示されてもよい。また、表示画面300の下部に音声認識用単語301がスクロール表示されてもよい。

[0134] また、メモリ13に記憶されている全ての音声認識用単語が表示されるのではなく、メモリ13に記憶されている音声認識用単語のうちの一部が表示されてもよい。また、ユーザの使用頻度が高い音声認識用単語が表示されてもよい。また、ユーザによって設定された音声認識用単語がメモリ13に記憶されてもよい。

[0135] また、メモリ13は、複数種類の機器の操作に関する音声認識用単語を記憶してもよく、音声認識用単語表示制御部212は、ユーザによって操作される機器に関する音声認識用単語をメモリ13から読み出し、表示部16に表示してもよい。なお、サーバ2が音声認識部を備える場合、ユーザによっ

て操作される機器を特定するための情報がサーバ２に送信され、当該情報に基づいて音声認識用単語がメモリから読み出される。

[0136] また、本実施の形態１１では、実施の形態２の音声認識部の構成に音声認識用単語表示制御部２１２を付加しているが、本開示は特にこれに限定されず、実施の形態１，３～１０の音声認識部の構成に音声認識用単語表示制御部２１２を付加してもよい。

[0137] （実施の形態１２）

図１９は、本開示の実施の形態１２における音声認識部の構成を示す図である。本実施の形態１２における音声認識部１１１は、音声取得部２０１、空間音圧分布検出部２０２、点音源検出部２０３、音声認識制御部２０４、音声認識処理部２０５、点音源位置特定部２０６及び輝度制御部２１３を備える。なお、機器１の構成は、実施の形態１と同じであり、制御部１２が音声認識部１１１を備える。実施の形態１２において、実施の形態１～実施の形態１１と同一の構成については同一の符号を付し、説明を省略する。

[0138] 輝度制御部２１３は、音声認識制御部２０４によって音声認識処理を行うと判断された場合の表示部１６の輝度を、音声認識制御部２０４によって音声認識処理を行わないと判断された場合の表示部１６の輝度よりも向上させる。

[0139] 図２０（Ａ）は、本開示の実施の形態１２において、音声認識処理が行われる場合に表示部に表示される表示画面の一例を示す図であり、図２０（Ｂ）は、本開示の実施の形態１２において、音声認識処理が行われない場合に表示部に表示される表示画面の一例を示す図である。なお、図２０（Ａ）及び図２０（Ｂ）は、機器１がテレビである場合に表示部１６に表示される表示画面３１１，３１２の一例を示している。また、図２０（Ａ）及び図２０（Ｂ）では、ハッチングの有無により、輝度の違いを表現している。

[0140] 図２０（Ａ）及び図２０（Ｂ）に示すように、音声認識制御部２０４によって音声認識処理を行うと判断された場合の表示部１６の輝度は、音声認識制御部２０４によって音声認識処理を行わないと判断された場合の表示部１

6の輝度よりも高くなる。

[0141] このように、音声認識処理が行われる場合の表示部16の輝度と、音声認識処理が行われない場合の表示部16の輝度とを異ならせることにより、音声認識処理が可能であるか否かをユーザに報知することができる。

[0142] なお、本実施の形態11では、輝度制御部213は、音声認識処理が行われる場合の表示画面の全体の輝度と、音声認識処理が行われない場合の表示画面の全体の輝度とを異ならせているが、本開示は特にこれに限定されず、輝度制御部213は、音声認識処理が行われる場合の表示画面の一部の輝度と、音声認識処理が行われない場合の表示画面の一部の輝度とを異ならせてもよい。例えば、輝度制御部213は、音声認識処理が行われる場合の表示画面の外縁部分の輝度と、音声認識処理が行われない場合の表示画面の外縁部分の輝度とを異ならせてもよい。

[0143] また、本実施の形態12では、実施の形態2の音声認識部の構成に輝度制御部213を付加しているが、本開示は特にこれに限定されず、実施の形態1、3～11の音声認識部の構成に輝度制御部213を付加してもよい。

[0144] (実施の形態13)

図21は、本開示の実施の形態13における音声認識部の構成を示す図である。本実施の形態13における音声認識部112は、音声取得部201、空間音圧分布検出部202、点音源検出部203、音声認識制御部204、音声認識処理部205、点音源位置特定部206、出力音声生成部208及びエージェント画像表示制御部214を備える。なお、機器1の構成は、実施の形態1と同じであり、制御部12が音声認識部111を備える。実施の形態12において、実施の形態1～実施の形態11と同一の構成については同一の符号を付し、説明を省略する。

[0145] 出力音声生成部208は、音声認識制御部204における音声認識処理を行うか否かの判断に応じた出力音声を生成し、スピーカ15から出力する。例えば、音声認識制御部204によって音声認識を行わないと判断された場合、出力音声生成部208は、ユーザに音声認識が可能な位置への移動を促

すような音声情報を生成する。

[0146] エージェント画像表示制御部 214 は、音声認識制御部 204 における音声認識処理を行うか否かの判断に応じたエージェント画像（アニメーション画像）を表示部 16 に表示する。例えば、音声認識制御部 204 によって音声認識を行わないと判断された場合、エージェント画像表示制御部 214 は、ユーザに音声認識が可能な位置への移動を促すようなエージェント画像を表示する。

[0147] エージェント画像表示制御部 214 は、出力音声生成部 208 によって生成された出力音声情報が出力される際に、出力される音声に連動してエージェント画像の表示を制御してもよい。また、メモリ 13 は、エージェント画像を予め記憶してもよく、エージェント画像表示制御部 214 は、音声認識制御部 204 における音声認識処理を行うか否かの判断に応じたエージェント画像をメモリ 13 から読み出し、表示部 16 に表示してもよい。

[0148] ここで、本開示の実施の形態 13 における音声認識システムの動作について説明する。図 22 は、本開示の実施の形態 13 における音声認識システムの動作を説明するためのシーケンス図である。

[0149] まず、ステップ S51 において、マイクロフォン 14 は、音声情報を取得する。

[0150] 次に、ステップ S52 において、マイクロフォン 14 は、取得した音声情報を音声認識部 112 へ出力する。

[0151] 次に、ステップ S53 において、音声取得部 201 は、マイクロフォン 14 によって出力された音声情報を取得する。

[0152] 次に、ステップ S54 において、空間音圧分布検出部 202 は、音声取得部 201 によって取得された音声情報に基づいて、空間内における音圧の分布を示す空間音圧分布を検出する。

[0153] 次に、ステップ S55 において、点音源検出部 203 は、空間音圧分布検出部 202 によって検出された空間音圧分布に基づいて、空間内における点音源を検出する。

- [0154] 次に、ステップS 5 6において、点音源位置特定部2 0 6は、点音源検出部2 0 3によって検出された点音源の空間内における位置を特定する。なお、点音源が検出されたと判断された場合、点音源位置特定部2 0 6は、点音源検出部2 0 3によって検出された点音源の空間内における位置を特定し、点音源が検出されないと判断された場合、点音源位置特定部2 0 6は、点音源検出部2 0 3によって検出された点音源の空間内における位置を特定しない。
- [0155] 次に、ステップS 5 7において、音声認識制御部2 0 4は、点音源の位置と機器1との距離が所定の距離より近いかなかを判断する。
- [0156] 次に、ステップS 5 8において、エージェント画像表示制御部2 1 4は、音声認識制御部2 0 4における音声認識処理を行うかなかの判断に応じたエージェント画像をメモリ1 3から読み出す。
- [0157] 次に、ステップS 5 9において、エージェント画像表示制御部2 1 4は、読み出したエージェント画像を表示部1 6へ出力する。
- [0158] 次に、ステップS 6 0において、表示部1 6は、エージェント画像表示制御部2 1 4によって出力されたエージェント画像を表示する。
- [0159] 次に、ステップS 6 1において、出力音声生成部2 0 8は、音声認識制御部2 0 4における音声認識処理を行うかなかの判断に応じた出力音声情報を生成する。
- [0160] 次に、ステップS 6 2において、出力音声生成部2 0 8は、生成した出力音声情報をスピーカ1 5へ出力する。
- [0161] 次に、ステップS 6 3において、スピーカ1 5は、出力音声生成部2 0 8によって出力された出力音声情報を出力する。
- [0162] 次に、ステップS 6 4において、音声認識処理部2 0 5は、音声認識制御部2 0 4によって音声認識処理を行うと判断された場合、音声取得部2 0 1によって取得された音声情報に対して音声認識処理を行う。また、音声認識処理部2 0 5は、音声認識制御部2 0 4によって音声認識処理を行わないと判断された場合、音声取得部2 0 1によって取得された音声情報に対して音

声認識処理を行わない。

[0163] 図 23 は、本開示の実施の形態 13 において、表示部に表示される表示画面の一例を示す図である。なお、図 23 は、機器 1 がテレビである場合に表示部 16 に表示される表示画面 400 の一例を示している。

[0164] 図 23 に示すように、エージェント画像表示制御部 214 は、人間の顔を模したエージェント画像 401 を表示部 16 の表示画面 400 に表示する。

[0165] エージェント画像 401 は、出力される音声に合わせて、あたかも話しているかのように動く。出力音声生成部 208 は、音声認識制御部 204 における音声認識処理を行うか否かの判断に応じた出力音声情報をメモリ 13 から読み出し、スピーカ 15 へ出力する。エージェント画像表示制御部 214 は、音声認識制御部 204 における音声認識処理を行うか否かの判断に応じたエージェント画像をメモリ 13 から読み出し、スピーカ 15 へ出力する。このとき、出力音声情報とエージェント画像とは対応付けて記憶されており、出力音声情報に連動したエージェント画像が表示される。

[0166] このように、音声認識制御部 204 における音声認識処理を行うか否かの判断に応じたエージェント画像（アニメーション画像）が表示部 16 に表示されるので、音声認識処理が可能であるか否かをユーザに報知することができる。

[0167] なお、本実施の形態 13 では、人間の顔を模したエージェント画像が表示されるが、本開示は特にこれに限定されない。

[0168] また、エージェント画像表示制御部 214 は、音声認識制御部 204 によって音声認識を行うと判断された場合、ユーザの音声操作に応じて動くエージェント画像を表示してもよい。

[0169] また、本実施の形態 13 では、表示画面 400 の下部にエージェント画像 401 が表示されているが、本開示は特にこれに限定されず、表示画面 400 の上部、中央部、左端部又は右端部にエージェント画像 401 が表示されてもよい。

[0170] また、本実施の形態 13 では、実施の形態 2 の音声認識部の構成に出力音

声生成部 208 及びエージェント画像表示制御部 214 を付加しているが、本開示は特にこれに限定されず、実施の形態 1, 3 ~ 12 の音声認識部の構成に出力音声生成部 208 及びエージェント画像表示制御部 214 を付加してもよい。

[0171] なお、上述した具体的実施形態には以下の構成を有する開示が主に含まれている。

[0172] 本開示の一局面に係る音声認識方法は、機器を動作させるための音声を認識する音声認識方法であって、所定の空間に配置されたマイクロフォンから音声情報を取得する音声情報取得ステップと、前記音声情報取得ステップにおいて取得された前記音声情報に基づいて、前記空間内における音圧の分布を示す空間音圧分布を検出する空間音圧分布検出ステップと、前記空間音圧分布検出ステップにおいて検出された前記空間音圧分布に基づいて、前記空間内における点音源を検出する点音源検出ステップと、前記点音源検出ステップにおいて前記点音源が検出された場合、前記音声情報取得ステップにおいて取得された前記音声情報に対して音声認識処理を行うと判断する音声認識制御ステップとを含む。

[0173] この構成によれば、点音源が検出された場合、取得された音声情報に対して音声認識処理が行われ、点音源が検出されない場合、音声認識処理が行われないので、音声認識処理を行うか否かを容易に判断することができるとともに、安価な構成で音声認識の誤動作を防止することができる。

[0174] また、上記の音声認識方法において、前記点音源検出ステップにおいて前記点音源が検出された場合、検出された前記点音源の前記空間内における位置を特定する位置特定ステップをさらに含み、前記音声認識制御ステップは、前記点音源の位置と前記機器との距離が所定の距離より近い場合、前記音声認識処理を行うと判断し、前記点音源の位置と前記機器との距離が所定の距離以上である場合、前記音声認識処理を行わないと判断することが好ましい。

[0175] この構成によれば、点音源の位置と機器との距離が所定の距離より近いか

否かを判断することにより、機器を操作するためにユーザが機器に近い場所で発話しているのか、又は機器の操作とは関係なくユーザが機器から離れた場所で発話しているのかを判断することができ、安価な構成で音声認識の誤動作を防止することができる。

[0176] また、上記の音声認識方法において、前記位置特定ステップにおいて特定された前記点音源の位置に向けて前記マイクロフォンの指向性の方向を制御する指向性制御ステップをさらに含むことが好ましい。

[0177] この構成によれば、点音源の位置に向けてマイクロフォンの指向性の方向を制御することにより、点音源からの音声情報を確実に取得することができ、音声認識の誤動作を防止することができる。

[0178] また、上記の音声認識方法において、スピーカから出力される出力音声情報を生成する出力音声生成ステップと、前記出力音声生成ステップにおいて生成された前記出力音声情報を出力音声として前記スピーカから出力する音声出力ステップと、前記出力音声生成ステップにおいて生成された前記出力音声情報を取得し、前記音声情報取得ステップにおいて取得された前記音声情報から、取得した前記出力音声情報をエコー成分として除去するエコー除去ステップとを含むことが好ましい。

[0179] この構成によれば、出力音声生成ステップにおいて生成された出力音声情報が取得され、音声情報取得ステップにおいて取得された音声情報から、取得した出力音声情報がエコー成分として除去されるので、点音源からの音声情報を確実に取得することができ、音声認識の誤動作を防止することができる。

[0180] また、上記の音声認識方法において、前記点音源検出ステップにおいて前記点音源が検出された場合、検出された前記点音源の前記空間内における位置を特定する位置特定ステップと、前記位置特定ステップにおいて位置が特定された前記点音源のうち、発せられた音声の前記マイクロフォンに直接到来する直接到来点音源を特定する直接到来点音源特定ステップをさらに含むことが好ましい。

[0181] この構成によれば、位置が特定された点音源のうち、発せられた音声が入力マイクに直接到来する直接到来点音源が特定されるので、機器を操作するためにユーザが機器の方向に向かって発話しているのか、又は機器の操作とは関係なくユーザが機器の方向とは異なる方向に向かって発話しているのかを判断することができる。

[0182] また、上記の音声認識方法において、前記音声認識制御ステップは、前記直接到来点音源特定ステップにおいて前記直接到来点音源が特定された場合、前記音声認識処理を行うと判断し、前記直接到来点音源特定ステップにおいて前記直接到来点音源が特定されない場合、前記音声認識処理を行わないと判断することが好ましい。

[0183] この構成によれば、発せられた音声が入力マイクに直接到来する場合、音声認識が行われ、発せられた音声が入力マイクに直接到来しない場合、音声認識が行われない。そのため、機器を操作するためにユーザが機器の方向に向かって発話しているのか、又は機器の操作とは関係なくユーザが機器の方向とは異なる方向に向かって発話しているのかを判断することができ、安価な構成で音声認識の誤動作を防止することができる。

[0184] また、上記の音声認識方法において、前記位置特定ステップにおいて位置が特定された前記点音源のうち、発せられた音声が入力マイクに直接到来する直接到来点音源を特定する直接到来点音源特定ステップをさらに含み、前記指向性制御ステップは、前記直接到来点音源特定ステップにおいて前記直接到来点音源が特定された場合、前記位置特定ステップにおいて特定された前記点音源の位置に向けて前記マイクの指向性の方向を制御することが好ましい。

[0185] この構成によれば、発せられた音声が入力マイクに直接到来する直接到来点音源の位置に向けてマイクの指向性の方向が制御されるので、点音源からの音声情報を確実に取得することができ、音声認識の誤動作を防止することができる。

[0186] また、上記の音声認識方法において、前記直接到来点音源特定ステップに

において前記直接到来点音源が特定された場合、前記直接到来した音声人間によって発話された音声であるか否かを識別する音声識別ステップをさらに含み、前記音声認識制御ステップは、前記音声識別ステップにおいて前記直接到来した音声人間によって発話された音声であると識別された場合、前記音声認識処理を行うと判断し、前記音声識別ステップにおいて前記直接到来した音声人間によって発話された音声ではないと識別された場合、前記音声認識処理を行わないと判断することが好ましい。

[0187] この構成によれば、点音源から直接到来した音声人間の音声である場合、音声認識が行われ、点音源から直接到来した音声人間の音声ではない場合、音声認識が行われないので、人間以外の機器から発せられる音声に対して音声認識処理が行われることを防止することができる。

[0188] また、上記の音声認識方法において、前記音声認識制御ステップにおいて前記音声認識処理を行うと判断された場合、前記音声認識処理に用いられる単語を予め記憶する記憶部から前記単語を読み出し、読み出した前記単語を表示部に表示する音声認識単語表示ステップをさらに含むことが好ましい。

[0189] この構成によれば、音声認識処理に用いられる音声認識用単語が表示されるので、ユーザは表示された音声認識用単語を確認し、確実に音声により機器を操作することができる。

[0190] また、上記の音声認識方法において、前記音声認識制御ステップにおいて前記音声認識処理を行うと判断された場合の表示部の輝度を、前記音声認識制御ステップにおいて前記音声認識処理を行わないと判断された場合の表示部の輝度よりも向上させる輝度制御ステップをさらに含むことが好ましい。

[0191] この構成によれば、音声認識処理が行われる場合の表示部の輝度と、音声認識処理が行われない場合の表示部の輝度とが異なるので、音声認識処理が可能であるか否かをユーザに報知することができる。

[0192] また、上記の音声認識方法において、前記音声認識制御ステップにおける前記音声認識処理を行うか否かの判断に応じたアニメーション画像を表示部に表示するアニメーション画像表示ステップをさらに含むことが好ましい。

[0193] この構成によれば、音声認識処理を行うか否かの判断に応じたアニメーション画像が表示部に表示されるので、音声認識処理が可能であるか否かをユーザに報知することができる。

[0194] 本開示の他の局面に係る音声認識装置は、機器を動作させるための音声を認識する音声認識装置であって、所定の空間に配置されたマイクロフォンから音声情報を取得する音声情報取得部と、前記音声情報取得部によって取得された前記音声情報に基づいて、前記空間内における音圧の分布を示す空間音圧分布を検出する空間音圧分布検出部と、前記空間音圧分布検出部によって検出された前記空間音圧分布に基づいて、前記空間内における点音源を検出する点音源検出部と、前記点音源検出部によって前記点音源が検出された場合、前記音声情報取得部によって取得された前記音声情報に対して音声認識処理を行うと判断する音声認識制御部とを備える。

[0195] この構成によれば、点音源が検出された場合、取得された音声情報に対して音声認識処理が行われ、点音源が検出されない場合、音声認識処理が行われないので、音声認識処理を行うか否かを容易に判断することができるとともに、安価な構成で音声認識の誤動作を防止することができる。

[0196] なお、開示を実施するための形態の項においてなされた具体的な実施態様または実施例は、あくまでも、本開示の技術内容を明らかにするものであって、そのような具体例にのみ限定して狭義に解釈されるべきものではなく、本開示の精神と特許請求事項との範囲内で、種々変更して実施することができるものである。

産業上の利用可能性

[0197] 本開示に係る音声認識方法及び音声認識装置は、安価な構成で音声認識の誤動作を防止することができ、機器を動作させるための音声を認識する音声認識方法及び音声認識装置として有用である。

符号の説明

- [0198] 1 機器
 1 1 通信部

- 1 2 制御部
- 1 3 メモリ
- 1 4 マイクロフォン
- 1 5 スピーカ
- 1 6 表示部
- 1 0 0 音声認識部
- 2 0 1 音声取得部
- 2 0 2 空間音圧分布検出部
- 2 0 3 点音源検出部
- 2 0 4 音声認識制御部
- 2 0 5 音声認識処理部

請求の範囲

- [請求項1] 機器を動作させるための音声を認識する音声認識方法であって、
 所定の空間に配置されたマイクロフォンから音声情報を取得する音声情報取得ステップと、
 前記音声情報取得ステップにおいて取得された前記音声情報に基づいて、前記空間内における点音源を検出する点音源検出ステップと、
 前記点音源検出ステップにおいて前記点音源が検出された場合、前記音声情報取得ステップにおいて取得された前記音声情報に対して音声認識処理を行うと判断する音声認識制御ステップとを含むことを特徴とする音声認識方法。
- [請求項2] 前記音声情報に基づいて前記空間内における音圧の分布を示す空間音圧分布を検出する空間音圧分布検出ステップをさらに含み、
 前記点音源検出ステップでは、前記空間音圧分布に基づいて前記点音源を検出する、
 請求項 1 に記載の音声認識方法。
- [請求項3] 前記空間音圧分布検出ステップにて検出する前記空間音圧分布は、所定角度ごとの音圧分布であり、
 前記点音源検出ステップでは前記音圧分布のうち、前記音声情報における音圧の大きさが所定の閾値よりも高い、一または複数の音源の角度範囲を特定し、
 特定した前記音源の角度範囲の幅が所定の幅以下である音源を、点音源として特定する、
 請求項 2 に記載の音声認識方法。
- [請求項4] 前記音声認識制御ステップは、前記空間音圧分布に基づいて、前記マイクロフォンから所定の角度内の領域において検出された音源に対しては音声認識処理を行うと判断し、前記マイクロフォンから所定の角度外の領域において検出された音源に対しては音声認識処理を行わないと判断する、

請求項 3 に記載の音声認識方法。

[請求項5] 前記点音源検出ステップにおいて前記点音源が検出された場合、検出された前記点音源の前記空間内における位置を特定する位置特定ステップをさらに含み、

前記音声認識制御ステップは、前記点音源の位置と前記機器との距離が所定の距離より近い場合、前記音声認識処理を行うと判断し、前記点音源の位置と前記機器との距離が所定の距離以上である場合、前記音声認識処理を行わないと判断することを特徴とする請求項 1 ～ 4 のいずれかに記載の音声認識方法。

[請求項6] 前記位置特定ステップにおいて特定された前記点音源の位置に向けて前記マイクロフォンの指向性の方向を制御する指向性制御ステップをさらに含むことを特徴とする請求項 5 記載の音声認識方法。

[請求項7] スピーカから出力される出力音声情報を生成する出力音声生成ステップと、

前記出力音声生成ステップにおいて生成された前記出力音声情報を出力音声として前記スピーカから出力する音声出力ステップと、

前記出力音声生成ステップにおいて生成された前記出力音声情報を取得し、前記音声情報取得ステップにおいて取得された前記音声情報から、取得した前記出力音声情報をエコー成分として除去するエコー除去ステップとを含むことを特徴とする請求項 1 ～ 6 のいずれかに記載の音声認識方法。

[請求項8] 前記点音源検出ステップにおいて前記点音源が検出された場合、検出された前記点音源の前記空間内における位置を特定する位置特定ステップと、

前記位置特定ステップにおいて位置が特定された前記点音源のうち、発せられた音声の前記マイクロフォンに直接到来する直接到来点音源を特定する直接到来点音源特定ステップをさらに含むことを特徴とする請求項 1 記載の音声認識方法。

- [請求項9] 前記音声認識制御ステップは、前記直接到来点音源特定ステップにおいて前記直接到来点音源が特定された場合、前記音声認識処理を行うと判断し、前記直接到来点音源特定ステップにおいて前記直接到来点音源が特定されない場合、前記音声認識処理を行わないと判断することを特徴とする請求項8記載の音声認識方法。
- [請求項10] 前記位置特定ステップにおいて位置が特定された前記点音源のうち、発せられた音声の前記マイクロフォンに直接到来する直接到来点音源を特定する直接到来点音源特定ステップをさらに含み、
前記指向性制御ステップは、前記直接到来点音源特定ステップにおいて前記直接到来点音源が特定された場合、前記位置特定ステップにおいて特定された前記点音源の位置に向けて前記マイクロフォンの指向性の方向を制御することを特徴とする請求項6記載の音声認識方法。
- [請求項11] 前記直接到来点音源特定ステップにおいて前記直接到来点音源が特定された場合、前記直接到来した音声人間によって発話された音声であるか否かを識別する音声識別ステップをさらに含み、
前記音声認識制御ステップは、前記音声識別ステップにおいて前記直接到来した音声人間によって発話された音声であると識別された場合、前記音声認識処理を行うと判断し、前記音声識別ステップにおいて前記直接到来した音声人間によって発話された音声ではないと識別された場合、前記音声認識処理を行わないと判断することを特徴とする請求項8記載の音声認識方法。
- [請求項12] 前記音声認識制御ステップにおいて前記音声認識処理を行うと判断された場合、前記音声認識処理に用いられる複数の単語を予め記憶する記憶部から、前記機器の操作に関する少なくとも1つの単語を読み出し、読み出した前記少なくとも1つの単語を表示部に表示する音声認識単語表示ステップをさらに含むことを特徴とする請求項1～11のいずれかに記載の音声認識方法。

- [請求項13] 前記音声認識制御ステップにおいて前記音声認識処理を行うと判断された場合の表示部の輝度を、前記音声認識制御ステップにおいて前記音声認識処理を行わないと判断された場合の表示部の輝度よりも向上させる輝度制御ステップをさらに含むことを特徴とする請求項1～12のいずれかに記載の音声認識方法。
- [請求項14] 前記音声認識制御ステップにおける前記音声認識処理を行うか否かの判断に応じたアニメーション画像を表示部に表示するアニメーション画像表示ステップをさらに含むことを特徴とする請求項1～13のいずれかに記載の音声認識方法。
- [請求項15] 機器を動作させるための音声を認識する音声認識装置であって、
所定の空間に配置されたマイクロフォンから音声情報を取得する音声情報取得部と、
前記音声情報取得部によって取得された前記音声情報に基づいて、前記空間内における点音源を検出する点音源検出部と、
前記点音源検出部によって前記点音源が検出された場合、前記音声情報取得部によって取得された前記音声情報に対して音声認識処理を行うと判断する音声認識制御部とを備えることを特徴とする音声認識装置。

補正された請求の範囲
[2014年12月19日(19.12.2014)国際事務局受理]

- [請求項1] (補正後) 機器を動作させるための音声を認識する音声認識方法であって、
所定の空間に配置されたマイクロフォンから音声情報を取得する音声情報取得ステップと、
前記音声情報取得ステップにおいて取得された前記音声情報に基づいて、前記空間内における点音源を検出する点音源検出ステップと、
前記点音源検出ステップにおいて前記点音源が検出された場合には前記音声情報取得ステップにおいて取得された前記音声情報に対して音声認識処理を行うと判断し、前記点音源が検出されなかった場合には前記音声情報取得ステップにおいて取得された前記音声情報に対して音声認識処理を行わないと判断する音声認識制御ステップと、を含むことを特徴とする音声認識方法。
- [請求項2] 前記音声情報に基づいて前記空間内における音圧の分布を示す空間音圧分布を検出する空間音圧分布検出ステップをさらに含み、
前記点音源検出ステップでは、前記空間音圧分布に基づいて前記点音源を検出する、
請求項1に記載の音声認識方法。
- [請求項3] 前記空間音圧分布検出ステップにて検出する前記空間音圧分布は、所定角度ごとの音圧分布であり、
前記点音源検出ステップでは前記音圧分布のうち、前記音声情報における音圧の大きさが所定の閾値よりも高い、一または複数の音源の角度範囲を特定し、
特定した前記音源の角度範囲の幅が所定の幅以下である音源を、点音源として特定する、
請求項2に記載の音声認識方法。
- [請求項4] 前記音声認識制御ステップは、前記空間音圧分布に基づいて、前記マイクロフォンから所定の角度内の領域において検出された音源に対しては音声認識処理を行うと判断し、前記マイクロフォンから所定の角度外の領域において検出された音源に対しては音声認識処理を行わないと判断する、
請求項3に記載の音声認識方法。
- [請求項5] 前記点音源検出ステップにおいて前記点音源が検出された場合、検出された前記点音源の前記空間内における位置を特定する位置特定ステップをさらに

含み、

前記音声認識制御ステップは、前記点音源の位置と前記機器との距離が所定の距離より近い場合、前記音声認識処理を行うと判断し、前記点音源の位置と前記機器との距離が所定の距離以上である場合、前記音声認識処理を行わないと判断することを特徴とする請求項 1～4 のいずれかに記載の音声認識方法。

[請求項6] 前記位置特定ステップにおいて特定された前記点音源の位置に向けて前記マイクロフォンの指向性の方向を制御する指向性制御ステップをさらに含むことを特徴とする請求項 5 記載の音声認識方法。

[請求項7] スピーカから出力される出力音声情報を生成する出力音声生成ステップと、前記出力音声生成ステップにおいて生成された前記出力音声情報を出力音声として前記スピーカから出力する音声出力ステップと、前記出力音声生成ステップにおいて生成された前記出力音声情報を取得し、前記音声情報取得ステップにおいて取得された前記音声情報から、取得した前記出力音声情報をエコー成分として除去するエコー除去ステップとを含むことを特徴とする請求項 1～6 のいずれかに記載の音声認識方法。

[請求項8] 前記点音源検出ステップにおいて前記点音源が検出された場合、検出された前記点音源の前記空間内における位置を特定する位置特定ステップと、前記位置特定ステップにおいて位置が特定された前記点音源のうち、発せられた音声の前記マイクロフォンに直接到来する直接到来点音源を特定する直接到来点音源特定ステップをさらに含むことを特徴とする請求項 1 記載の音声認識方法。

[請求項9] 前記音声認識制御ステップは、前記直接到来点音源特定ステップにおいて前記直接到来点音源が特定された場合、前記音声認識処理を行うと判断し、前記直接到来点音源特定ステップにおいて前記直接到来点音源が特定されない場合、前記音声認識処理を行わないと判断することを特徴とする請求項 8 記載の音声認識方法。

[請求項10] 前記位置特定ステップにおいて位置が特定された前記点音源のうち、発せられた音声の前記マイクロフォンに直接到来する直接到来点音源を特定する直接到来点音源特定ステップをさらに含み、

前記指向性制御ステップは、前記直接到来点音源特定ステップにおいて前記直接到来点音源が特定された場合、前記位置特定ステップにおいて特定された前記点音源の位置に向けて前記マイクロフォンの指向性の方向を制御することを特徴とする請求項6記載の音声認識方法。

[請求項11] 前記直接到来点音源特定ステップにおいて前記直接到来点音源が特定された場合、前記直接到来した音声人間によって発話された音声であるか否かを識別する音声識別ステップをさらに含み、
前記音声認識制御ステップは、前記音声識別ステップにおいて前記直接到来した音声人間によって発話された音声であると識別された場合、前記音声認識処理を行うと判断し、前記音声識別ステップにおいて前記直接到来した音声人間によって発話された音声ではないと識別された場合、前記音声認識処理を行わないと判断することを特徴とする請求項8記載の音声認識方法。

[請求項12] 前記音声認識制御ステップにおいて前記音声認識処理を行うと判断された場合、前記音声認識処理に用いられる複数の単語を予め記憶する記憶部から、前記機器の操作に関する少なくとも1つの単語を読み出し、読み出した前記少なくとも1つの単語を表示部に表示する音声認識単語表示ステップをさらに含むことを特徴とする請求項1～11のいずれかに記載の音声認識方法。

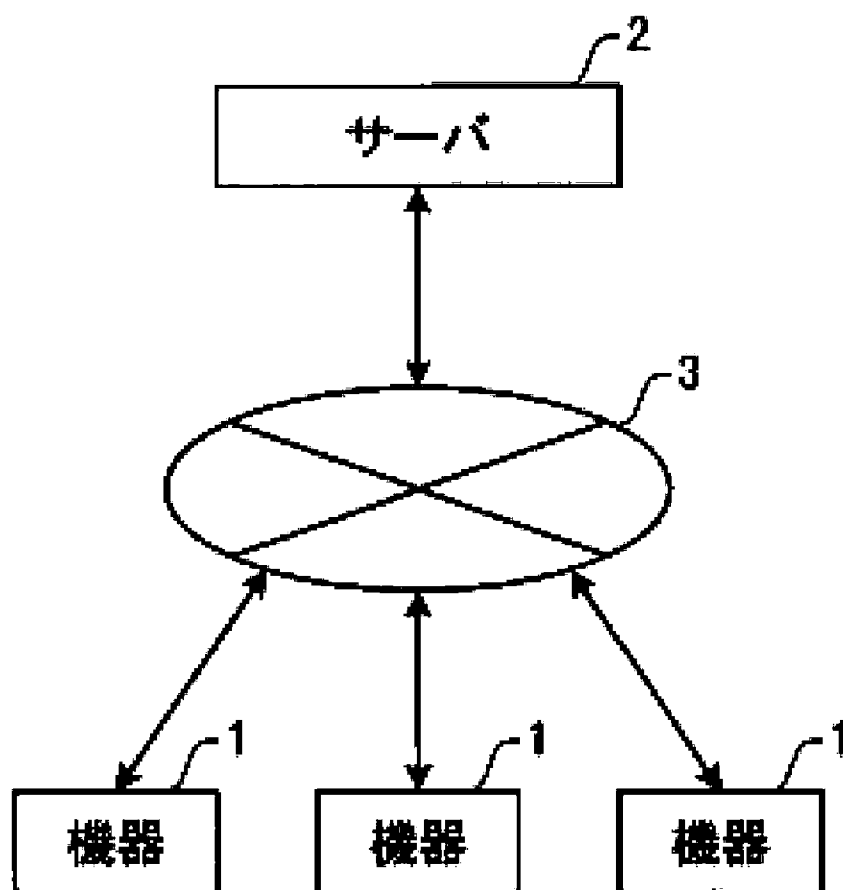
[請求項13] 前記音声認識制御ステップにおいて前記音声認識処理を行うと判断された場合の表示部の輝度を、前記音声認識制御ステップにおいて前記音声認識処理を行わないと判断された場合の表示部の輝度よりも向上させる輝度制御ステップをさらに含むことを特徴とする請求項1～12のいずれかに記載の音声認識方法。

[請求項14] 前記音声認識制御ステップにおける前記音声認識処理を行うか否かの判断に応じたアニメーション画像を表示部に表示するアニメーション画像表示ステップをさらに含むことを特徴とする請求項1～13のいずれかに記載の音声認識方法。

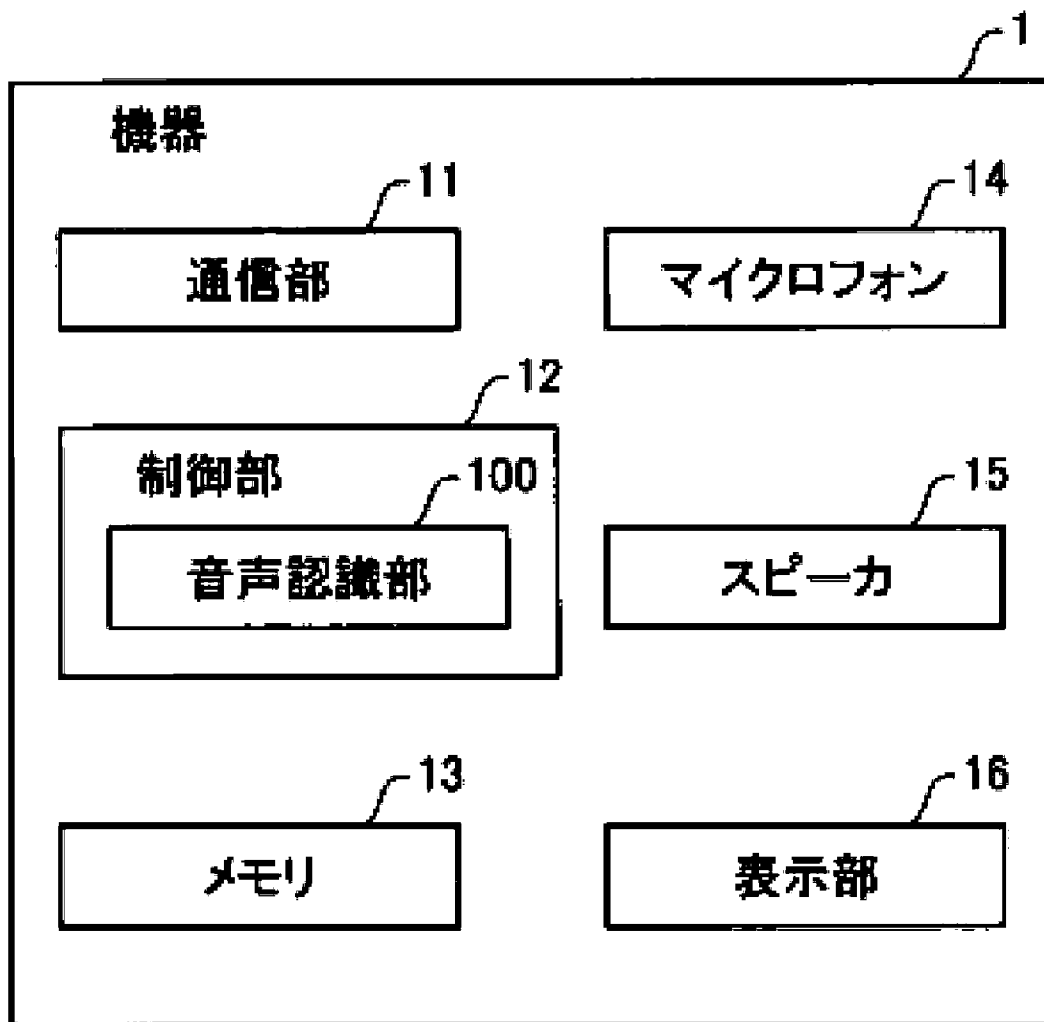
[請求項15] 機器を動作させるための音声を認識する音声認識装置であって、
所定の空間に配置されたマイクロフォンから音声情報を取得する音声情報取得部と、
前記音声情報取得部によって取得された前記音声情報に基づいて、前記空間

内における点音源を検出する点音源検出部と、
前記点音源検出部によって前記点音源が検出された場合、前記音声情報取得部によって取得された前記音声情報に対して音声認識処理を行うと判断する音声認識制御部とを備えることを特徴とする音声認識装置。

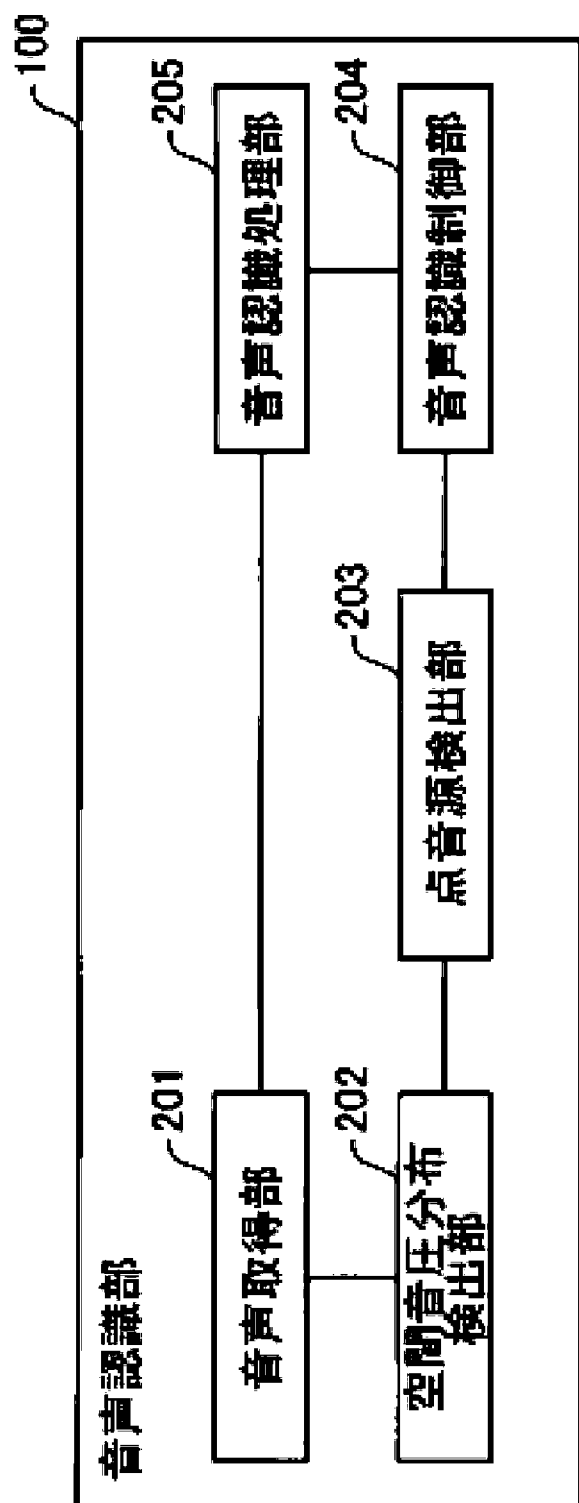
[図1]



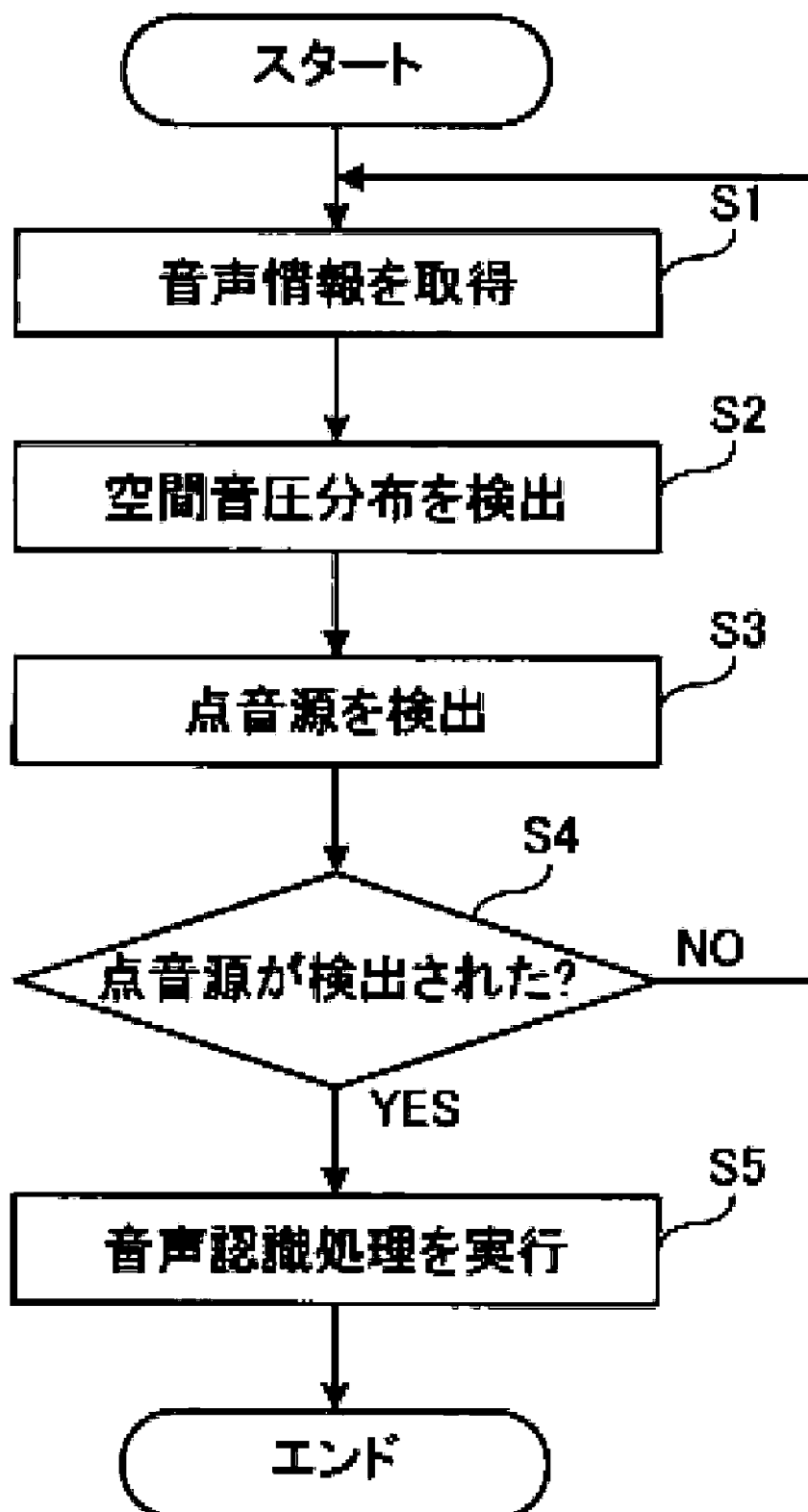
[図2]



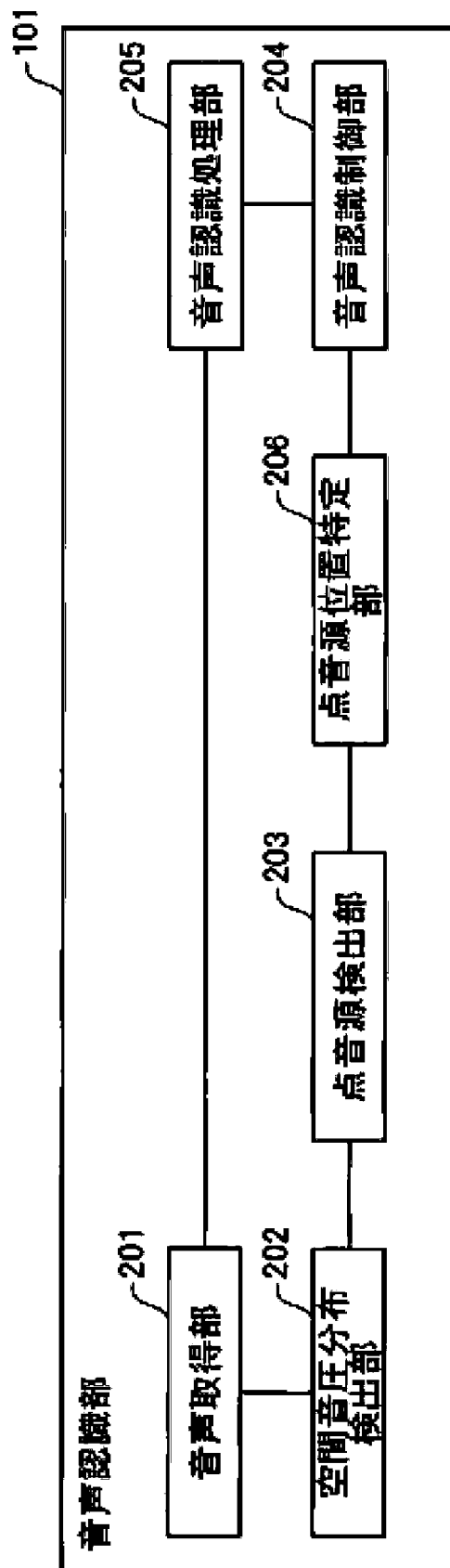
[図3]



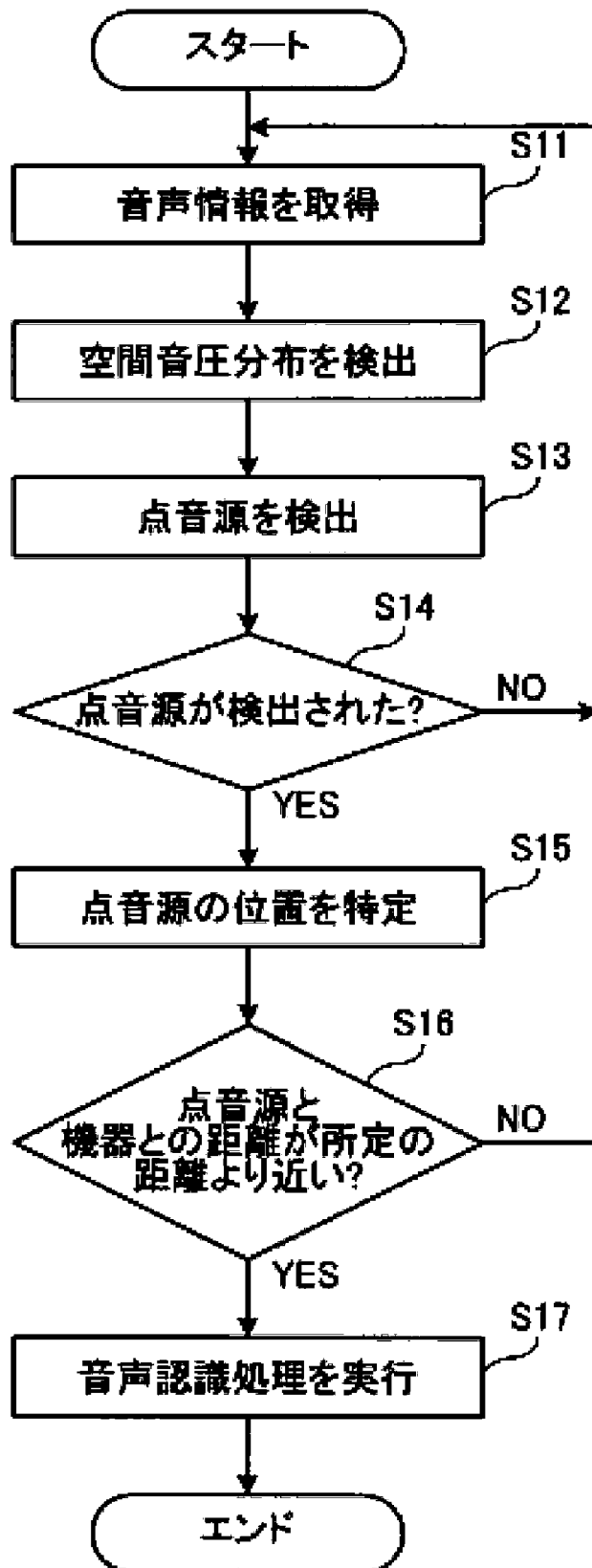
[図4]



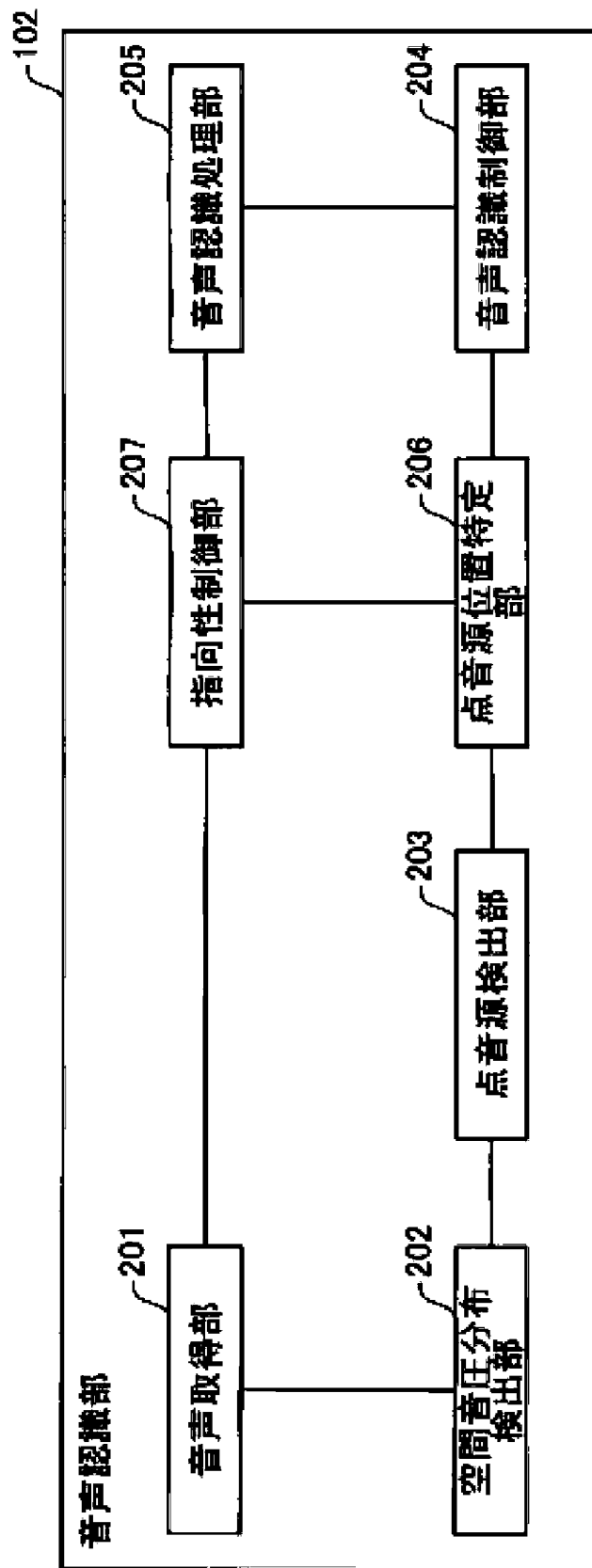
[図5]



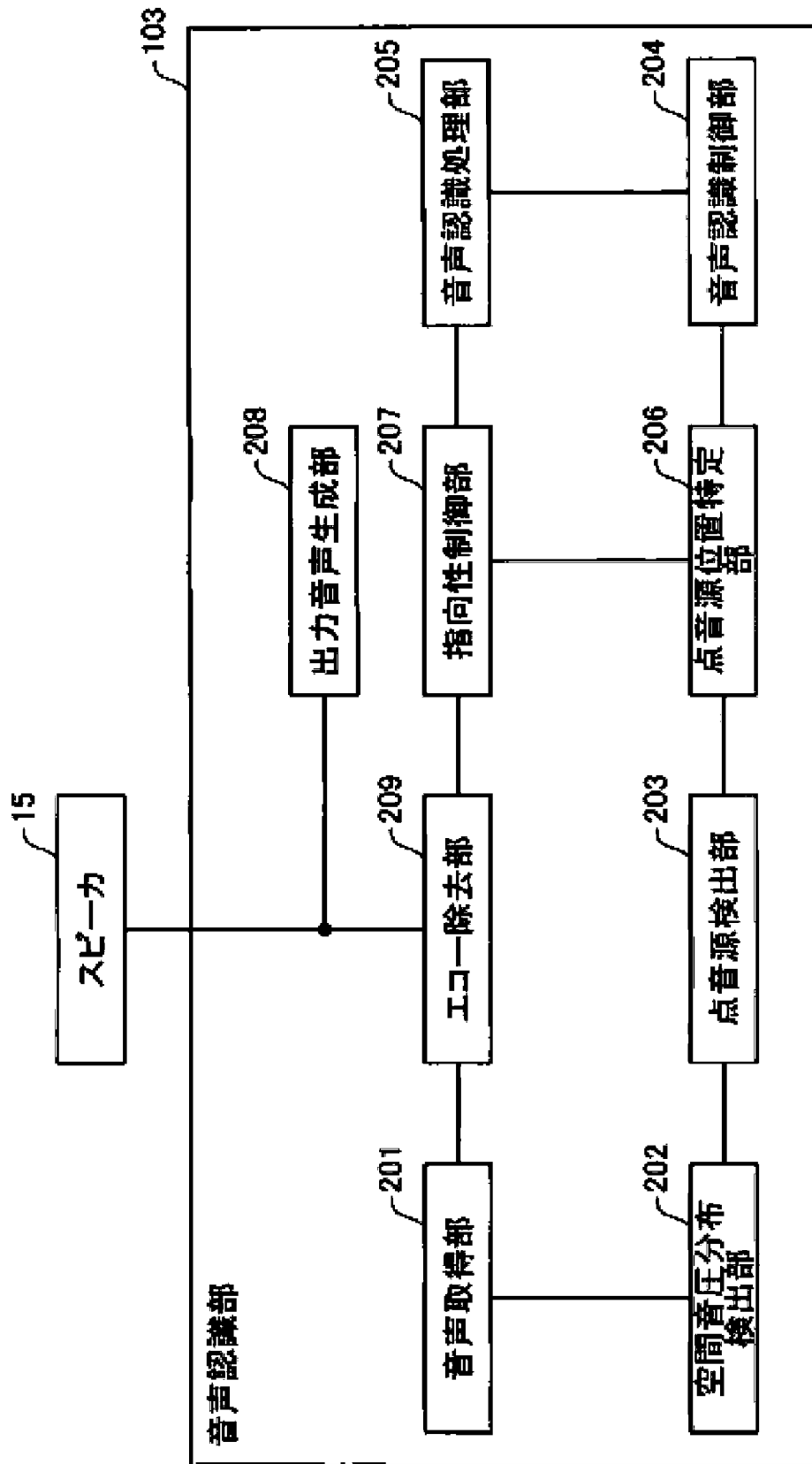
[図6]



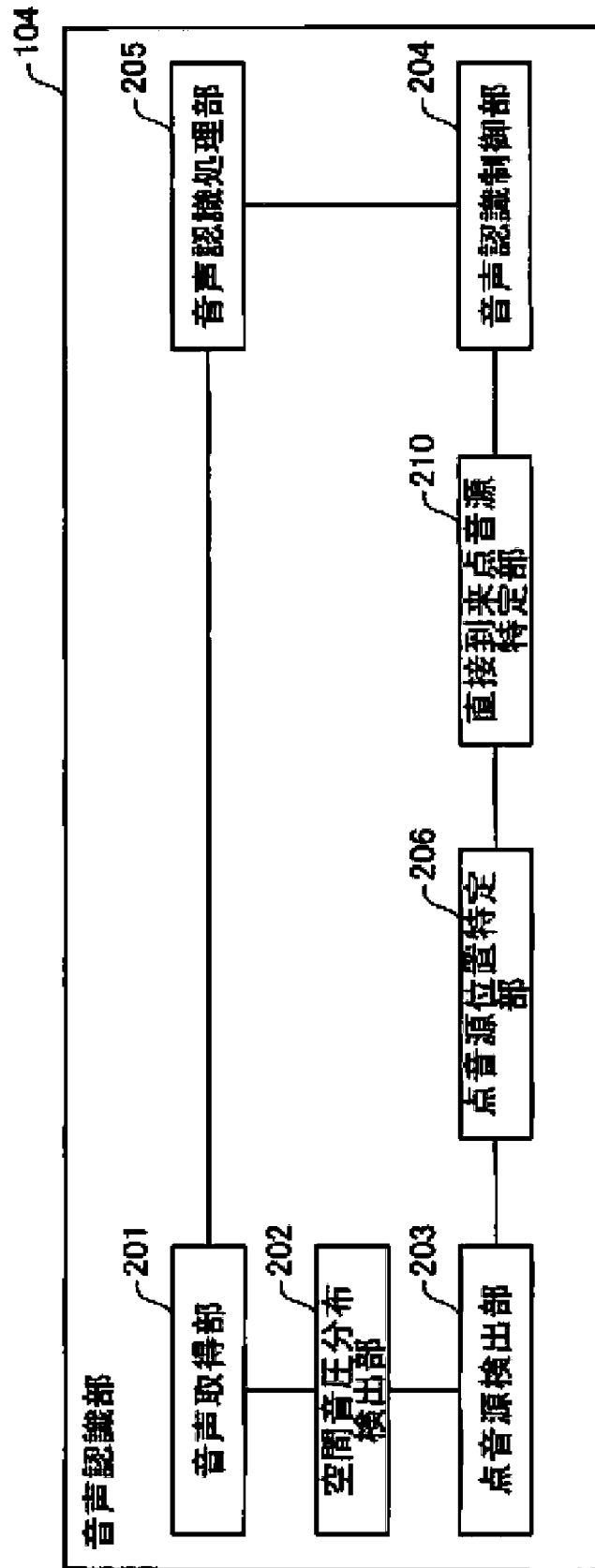
[図7]



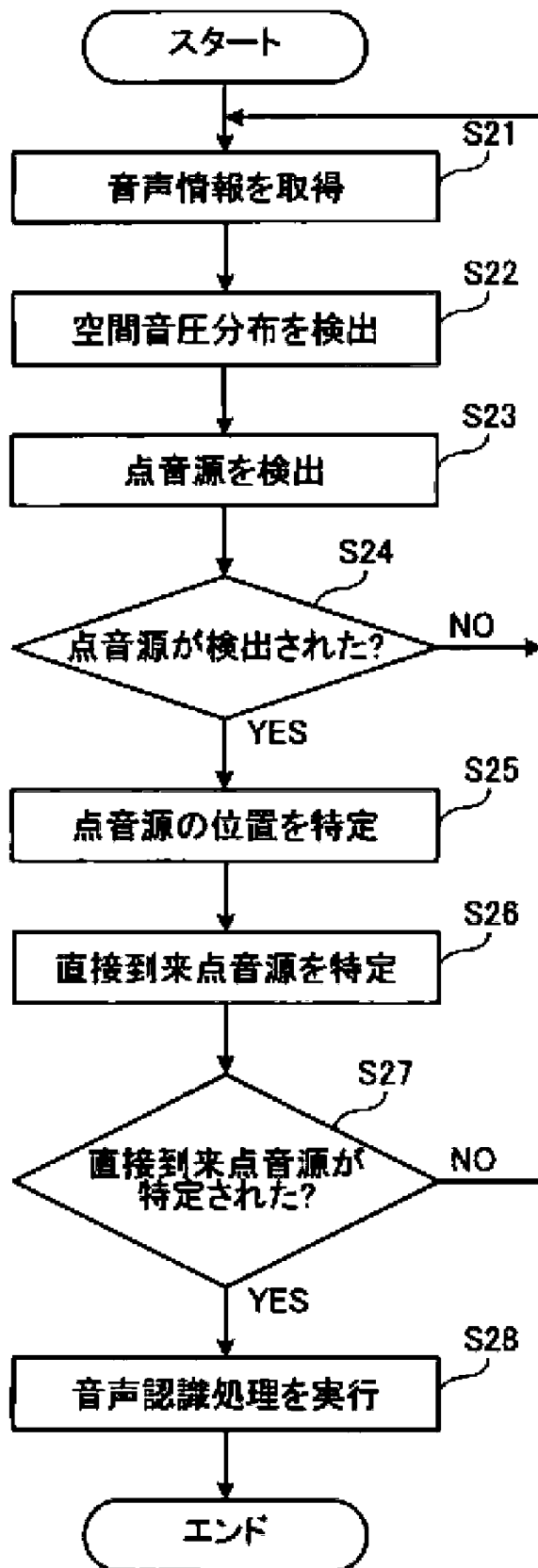
[図8]



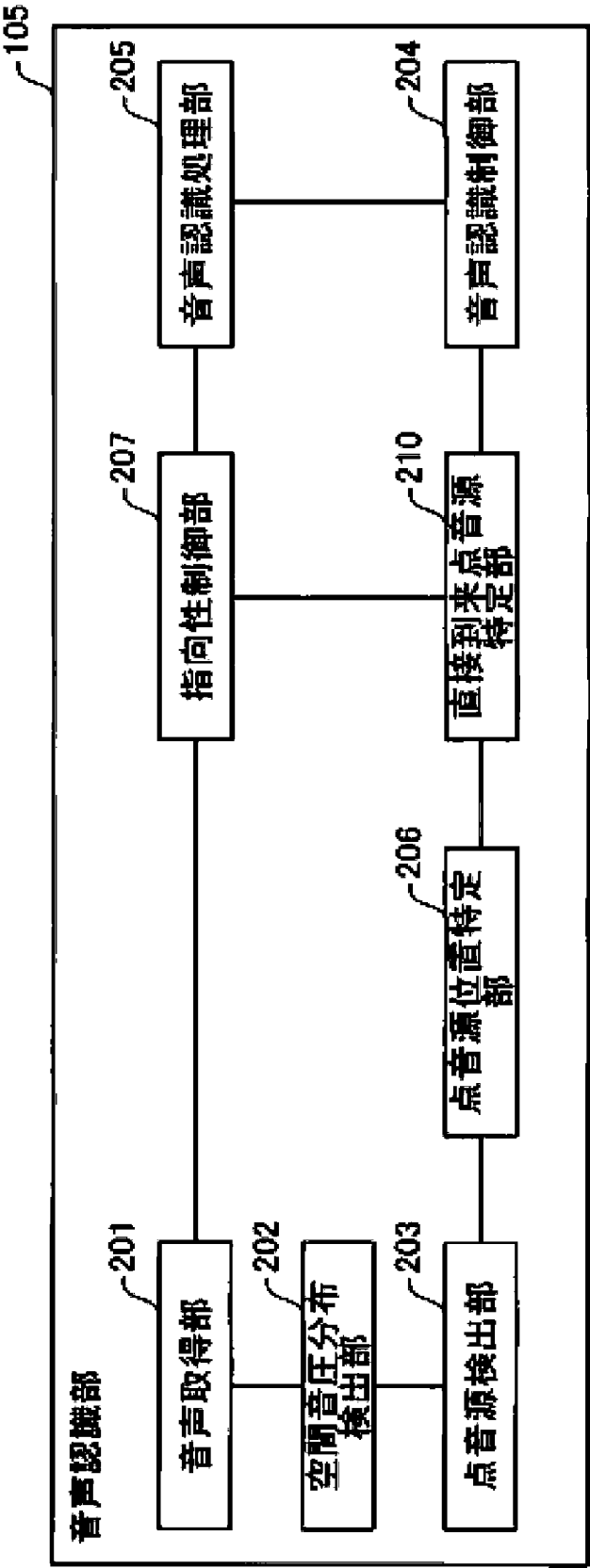
[図9]



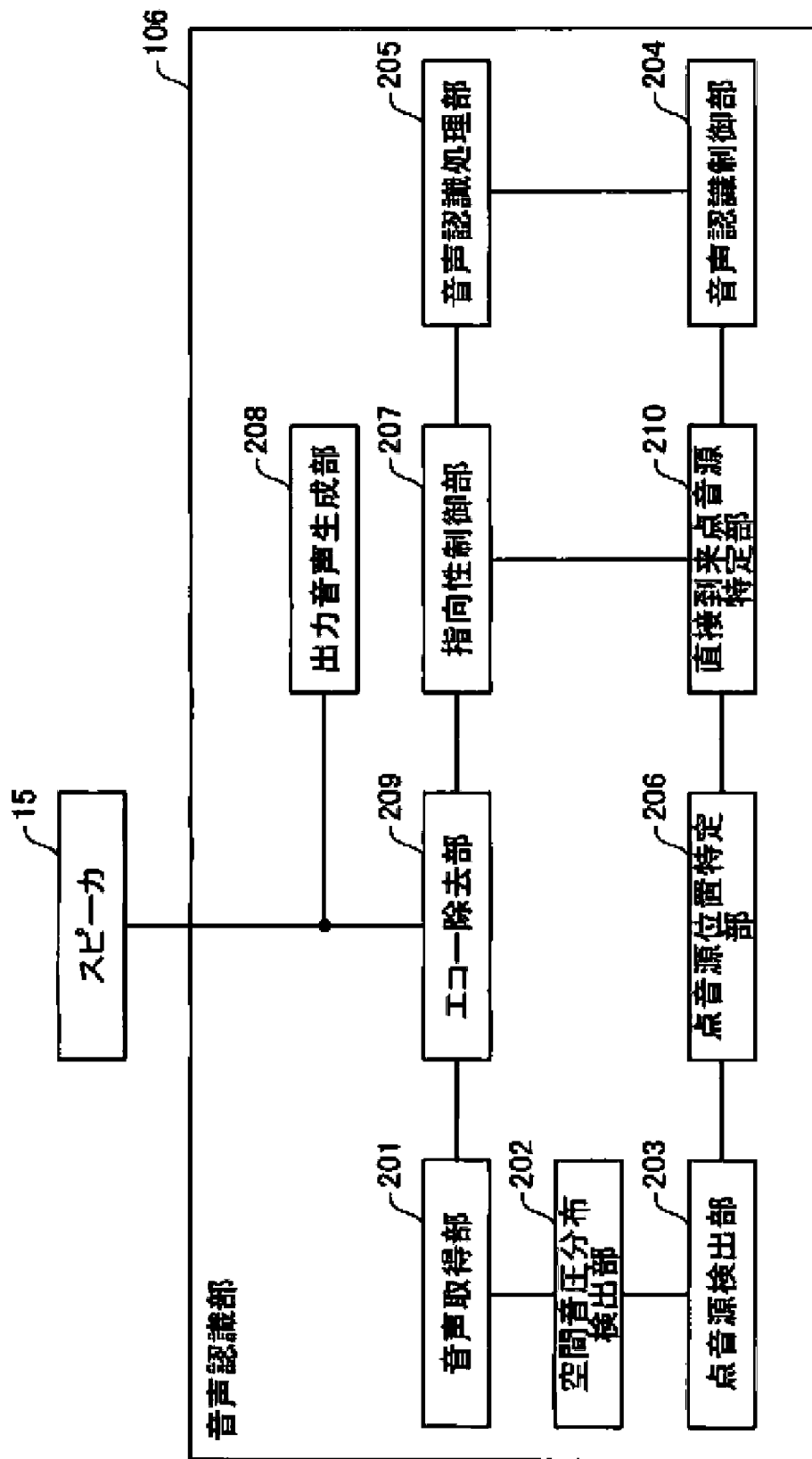
[図10]



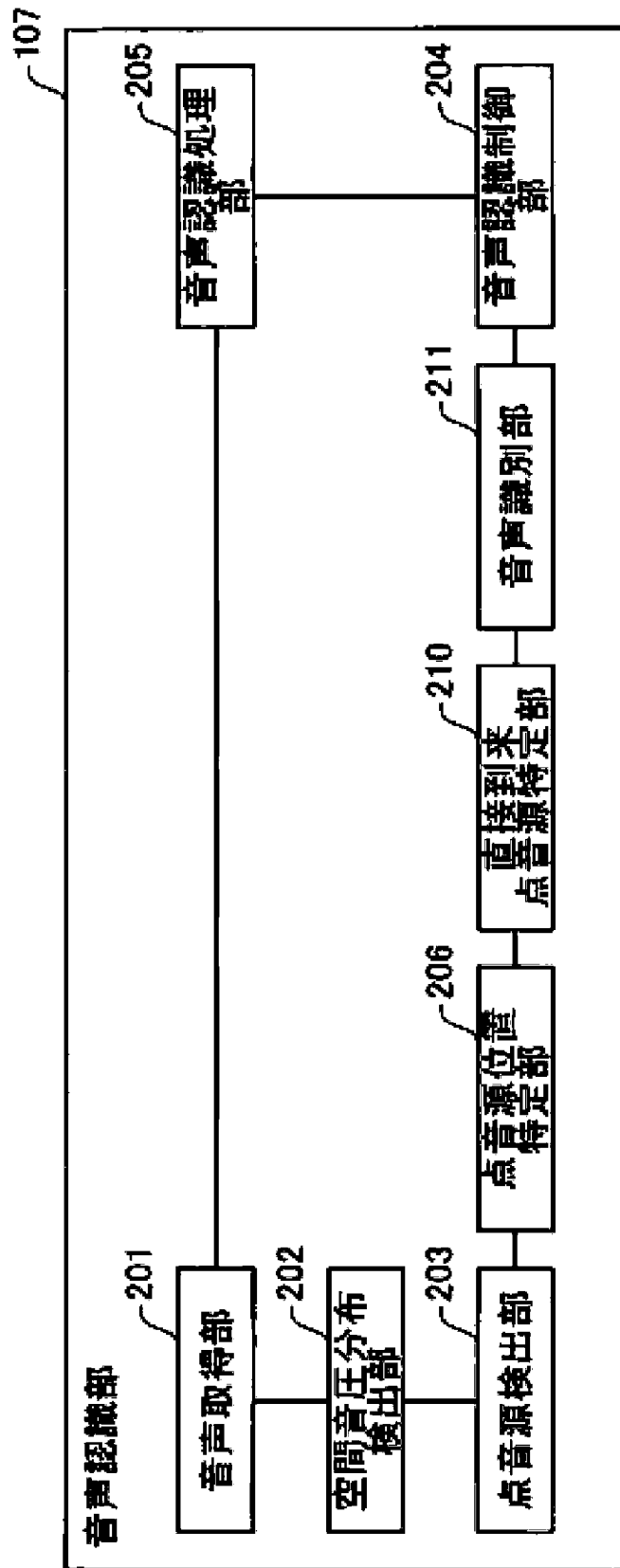
[図11]



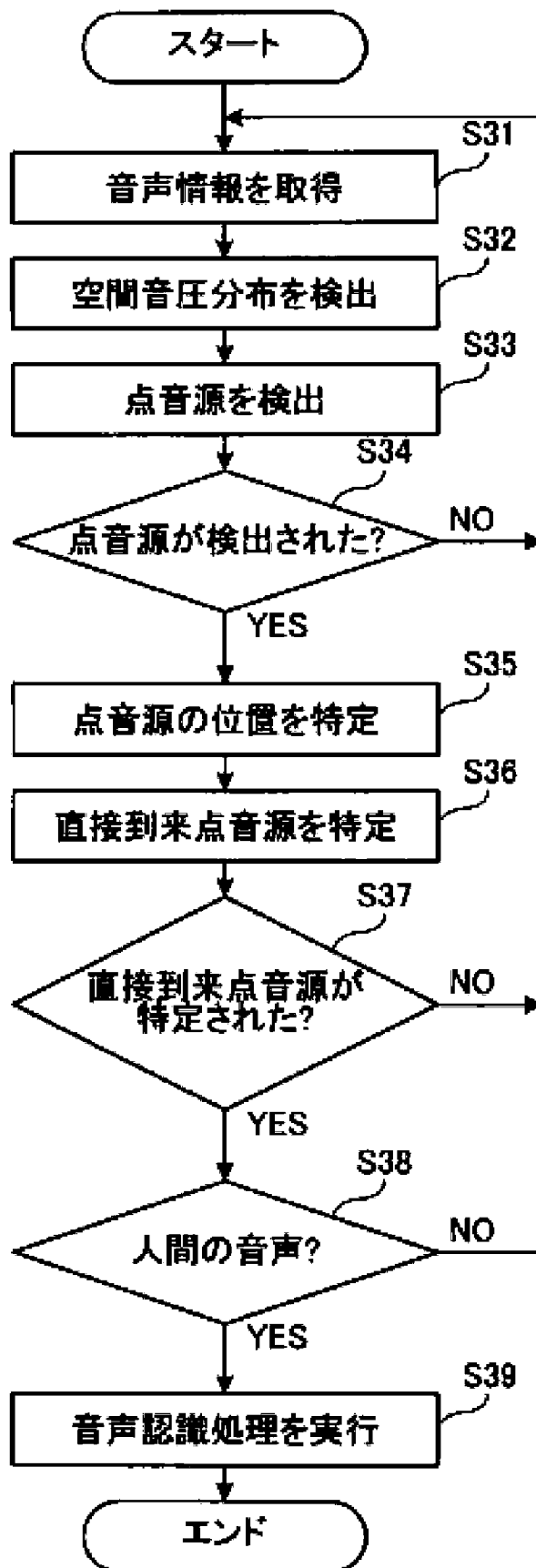
[図12]



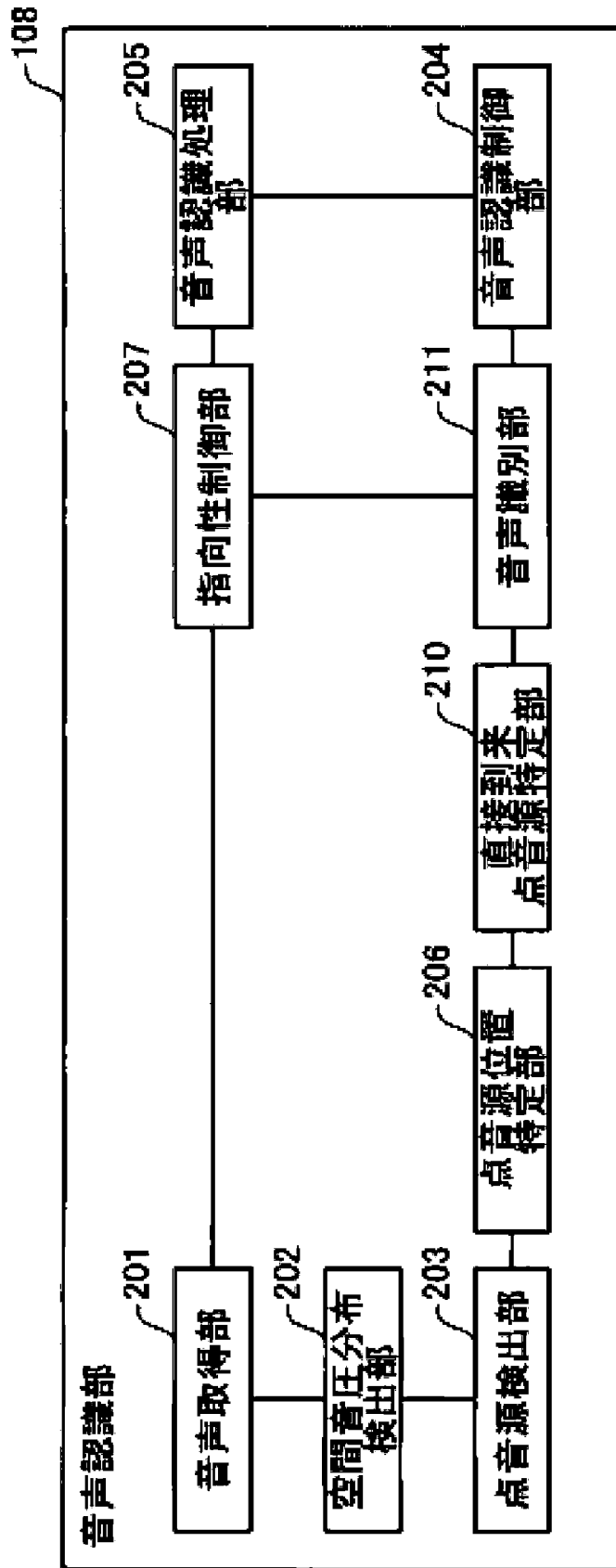
[図13]



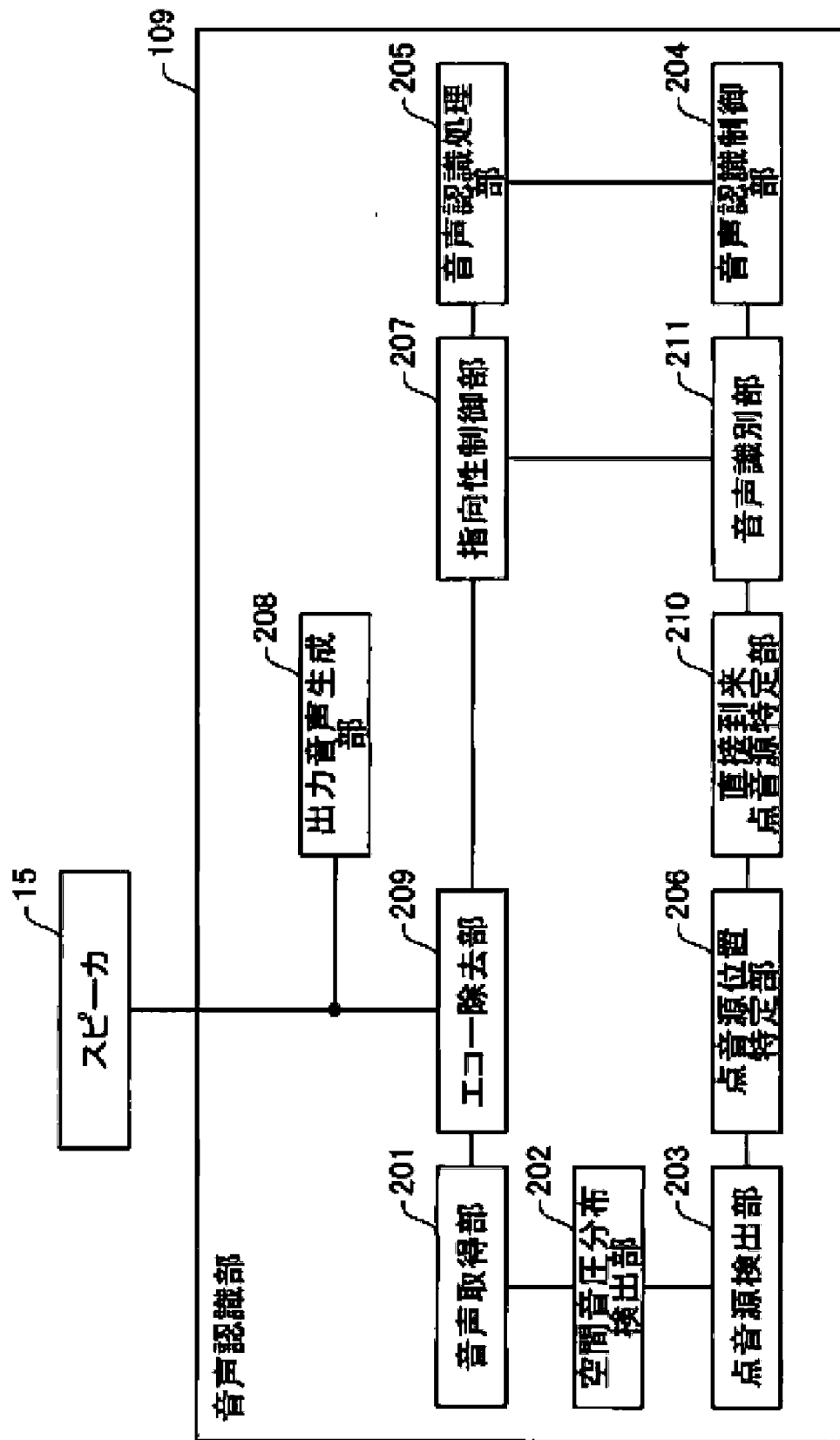
[図14]



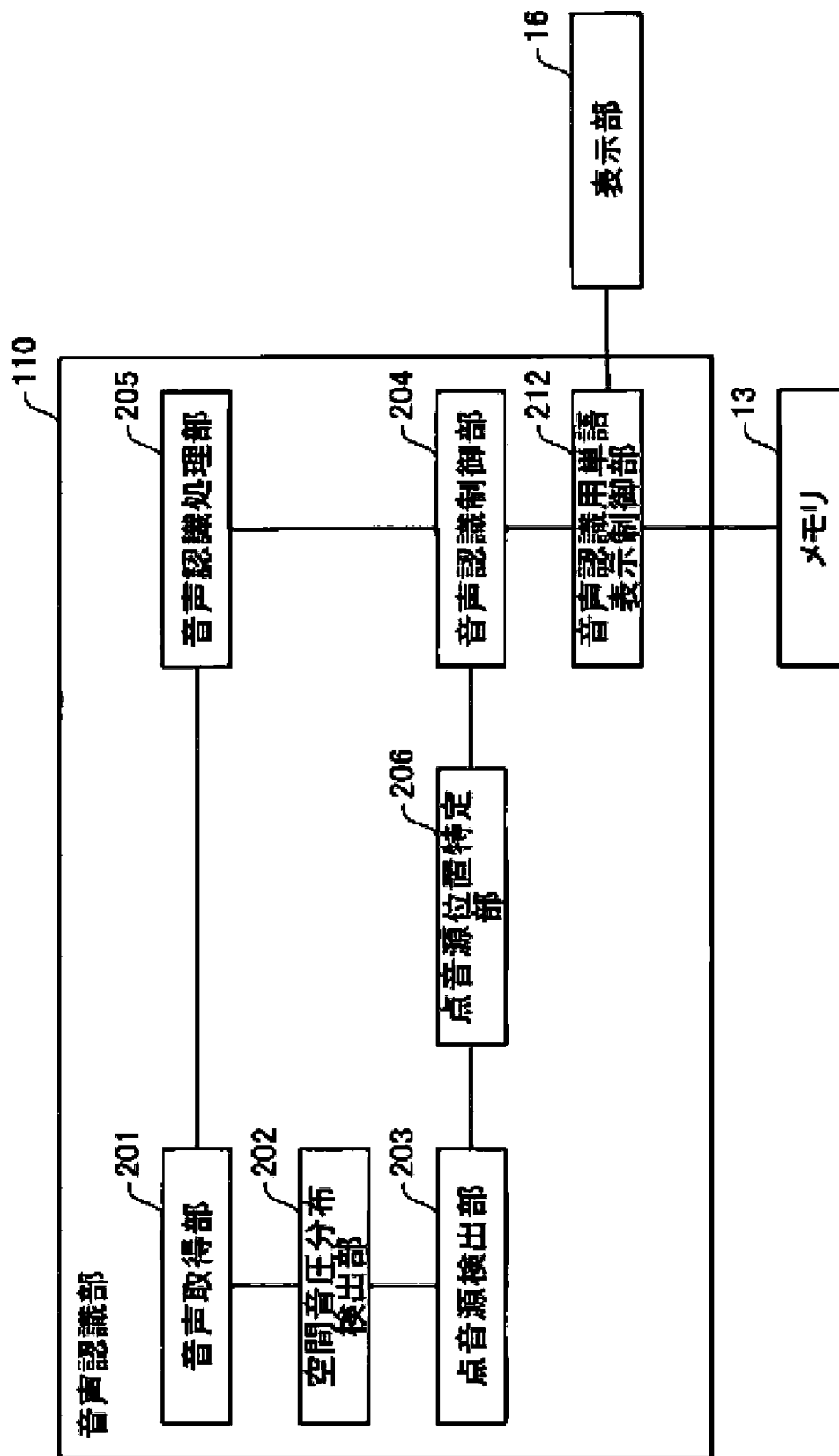
[図15]



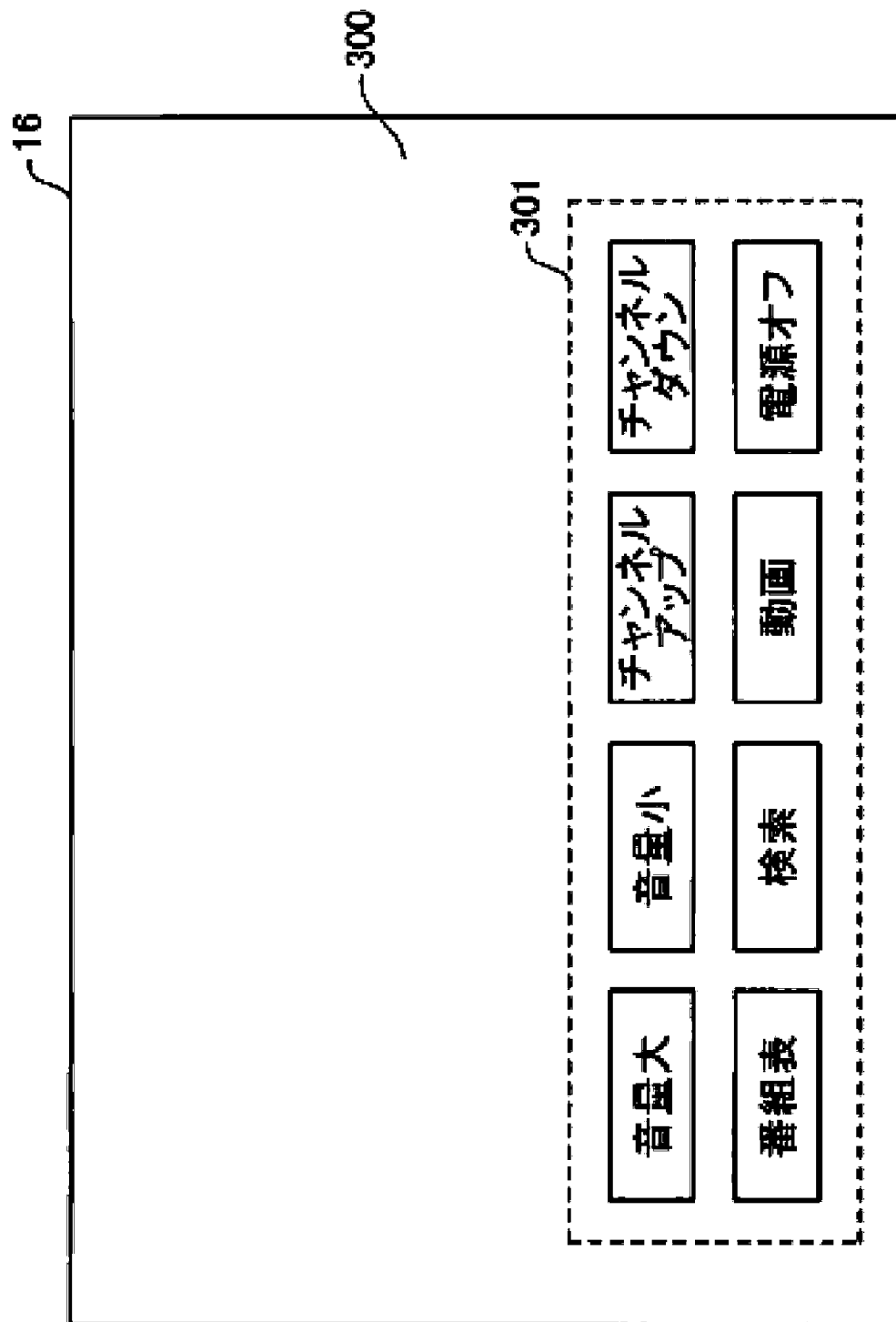
[図16]



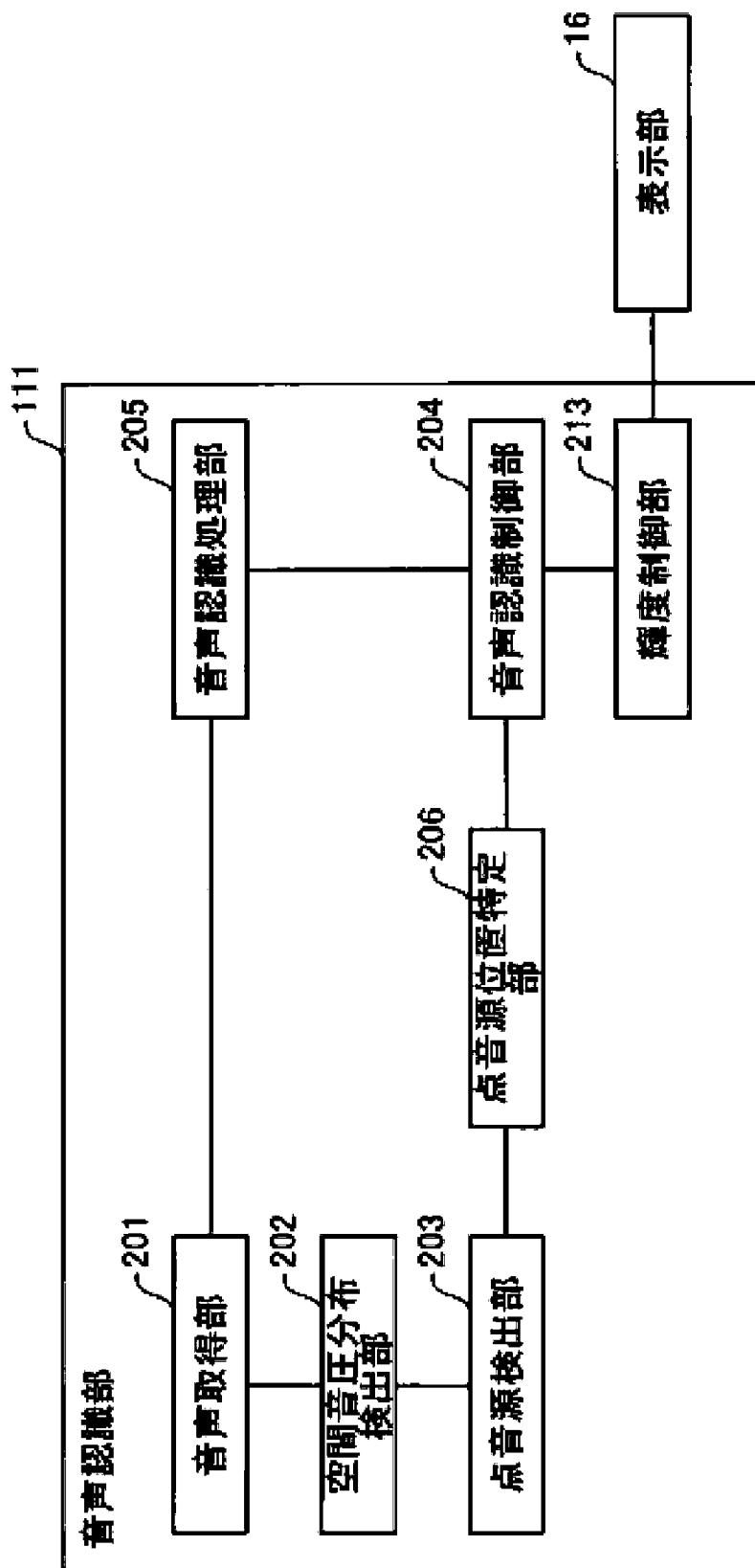
[図17]



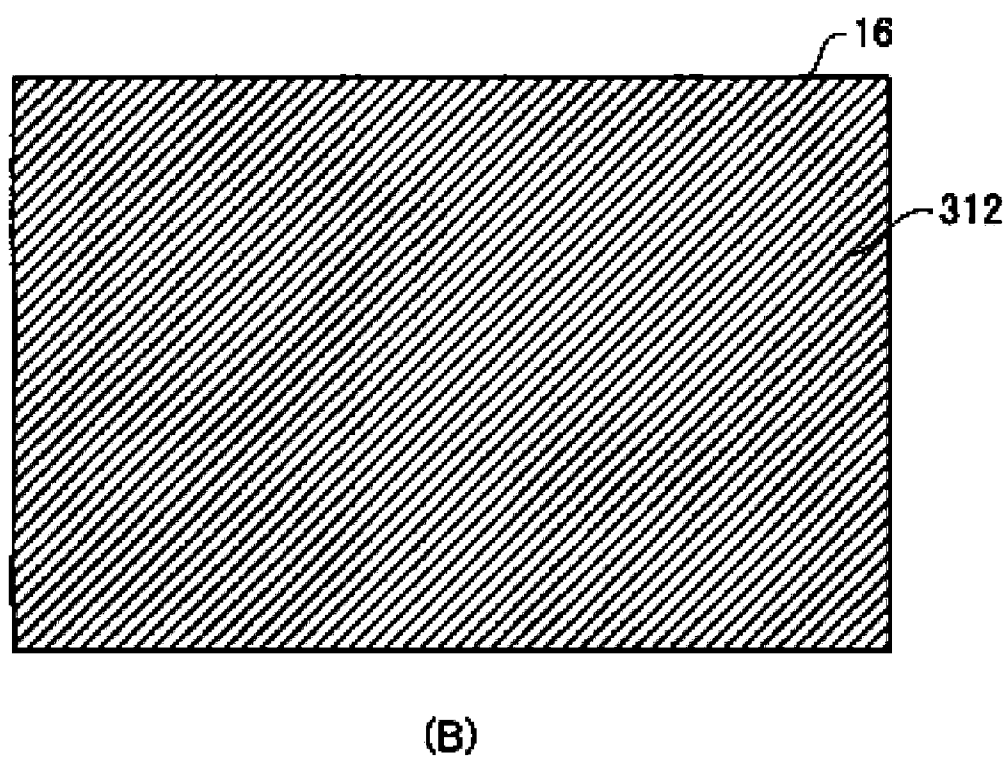
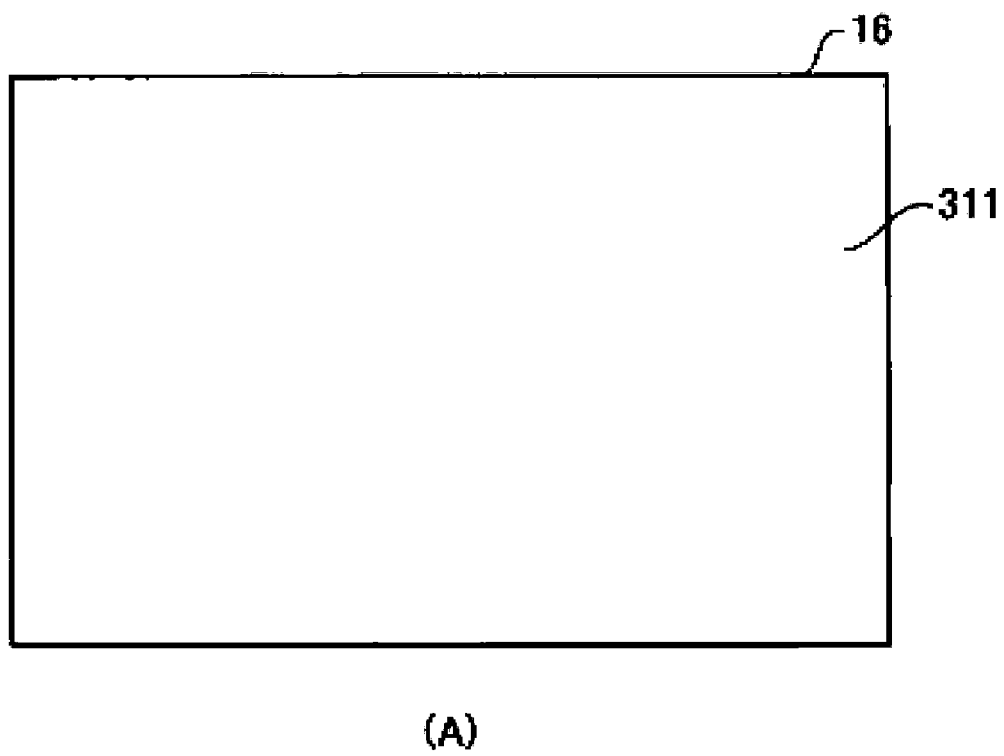
[図18]



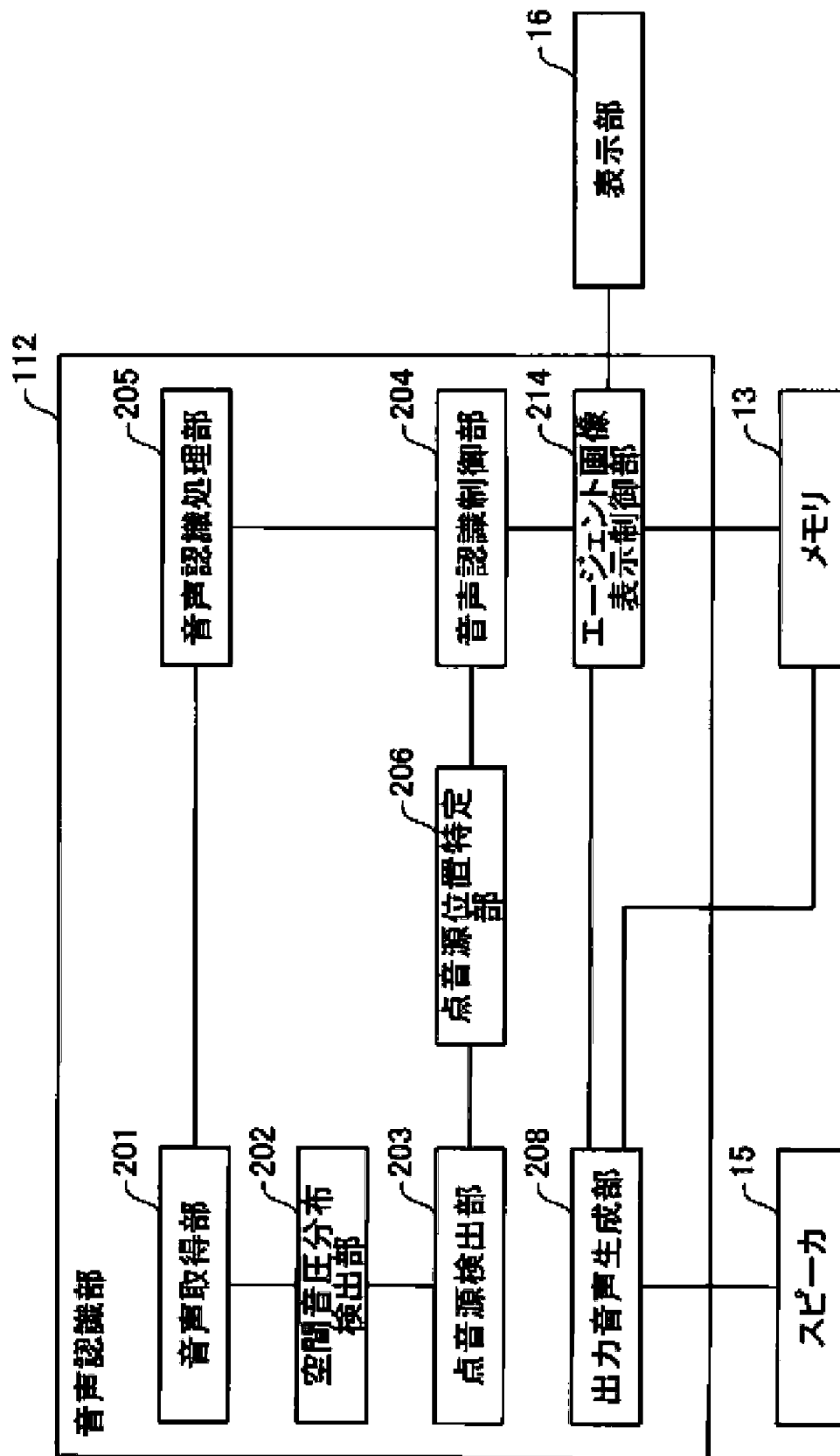
[図19]



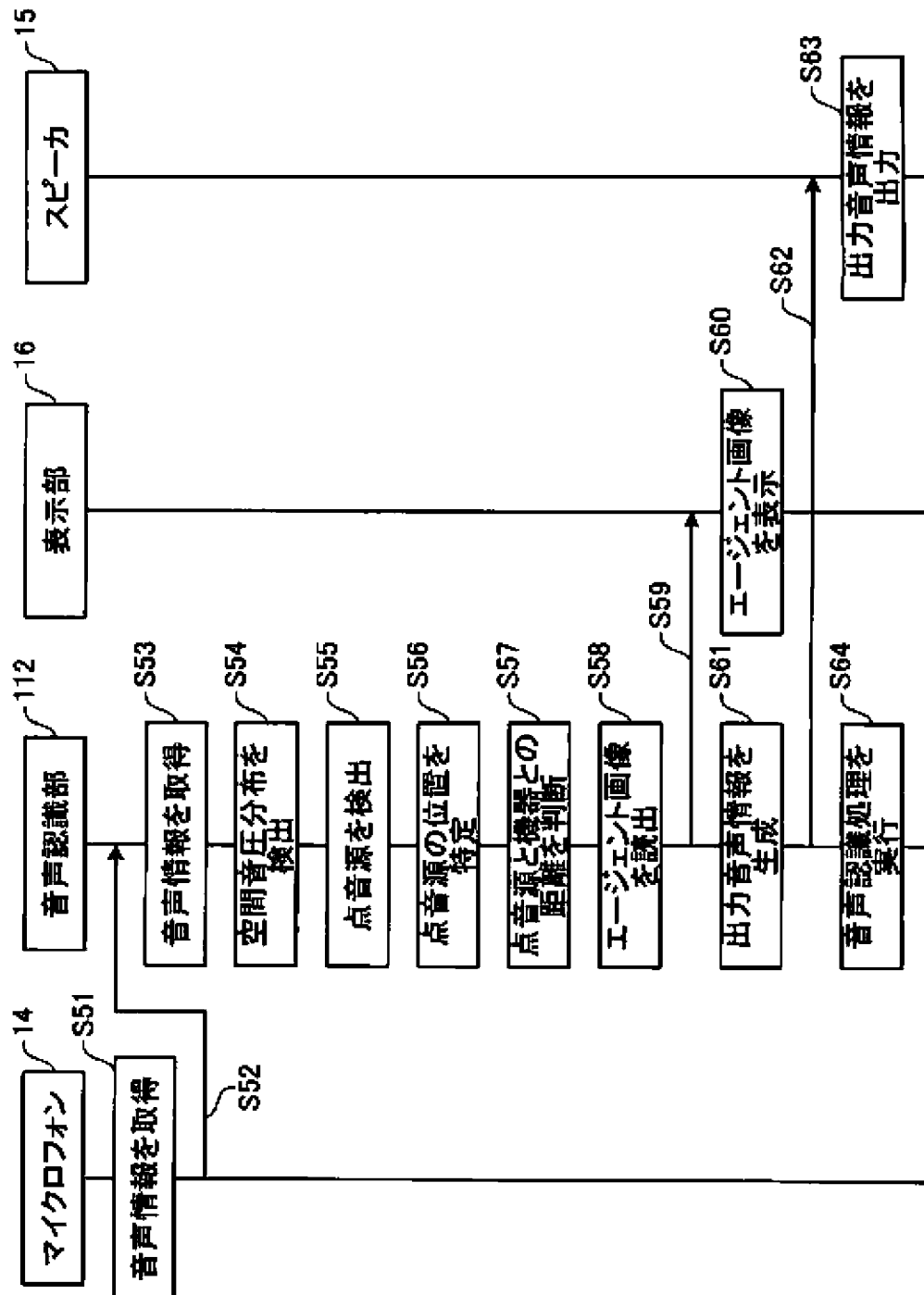
[図20]



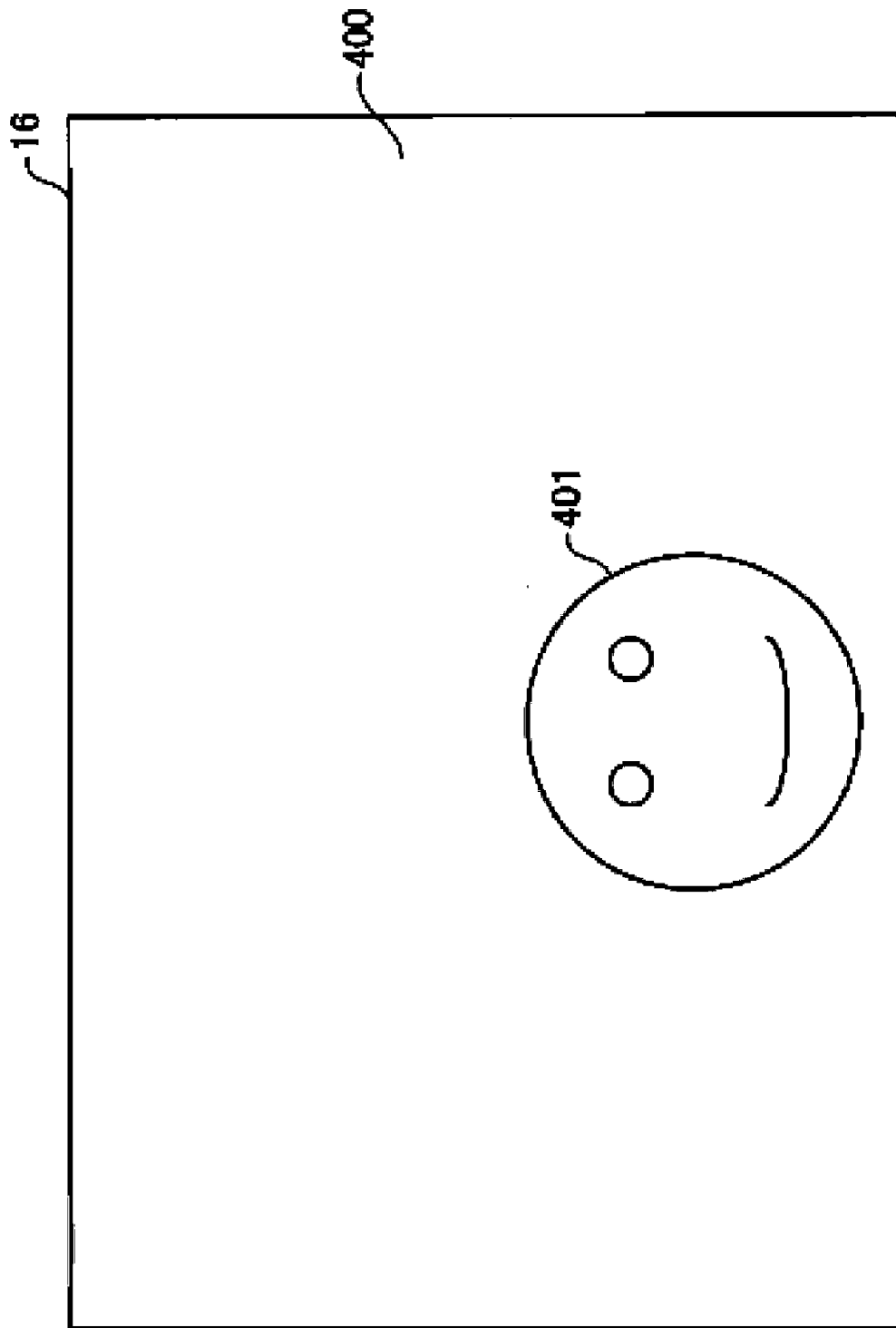
[図21]



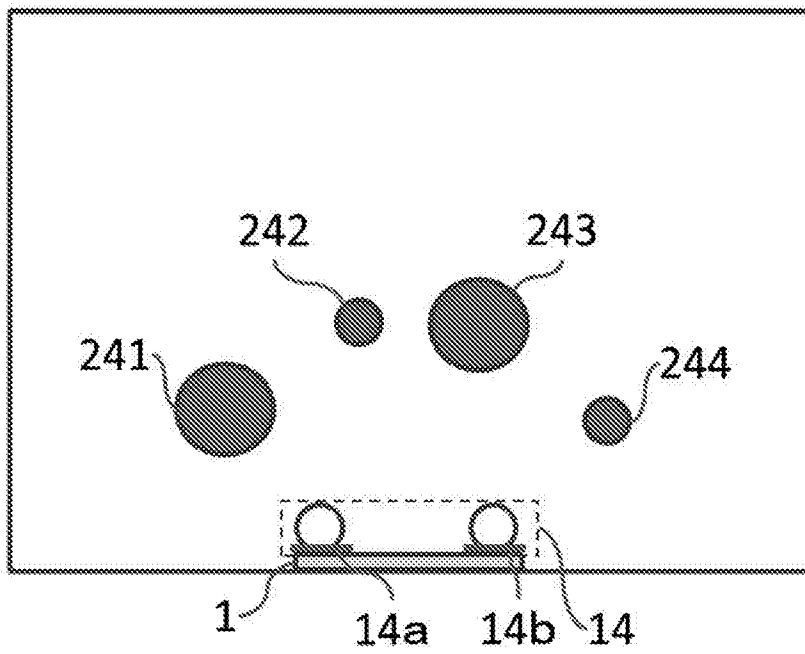
[図22]



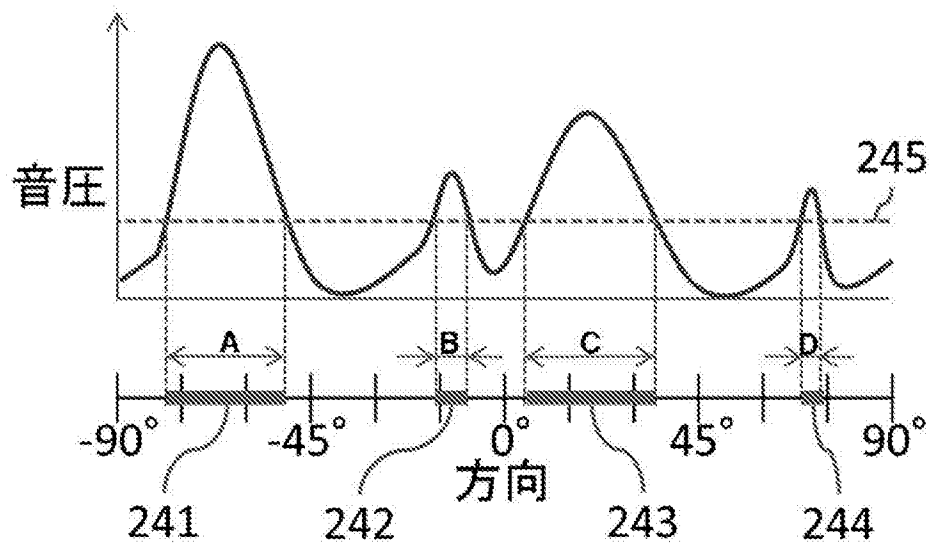
[図23]



[図24]

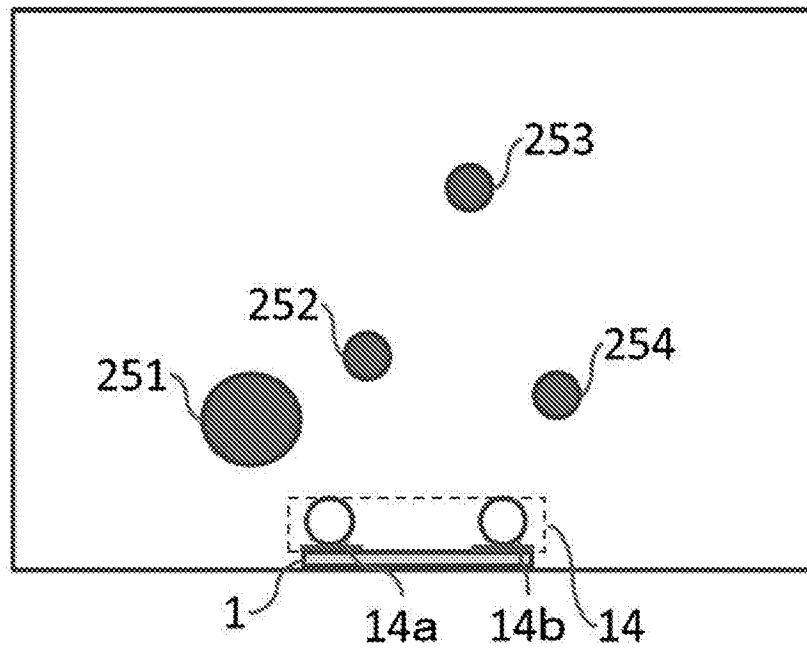


(A)

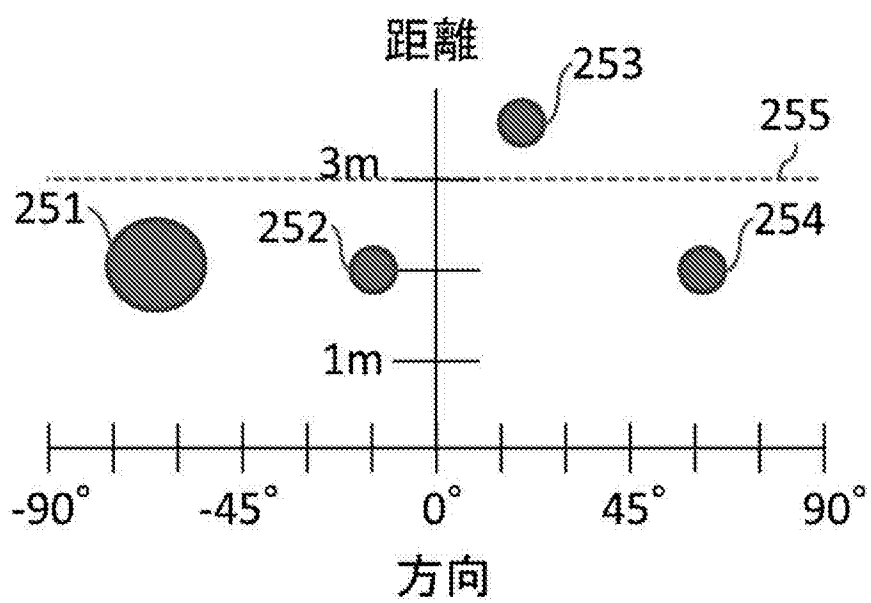


(B)

[図25]

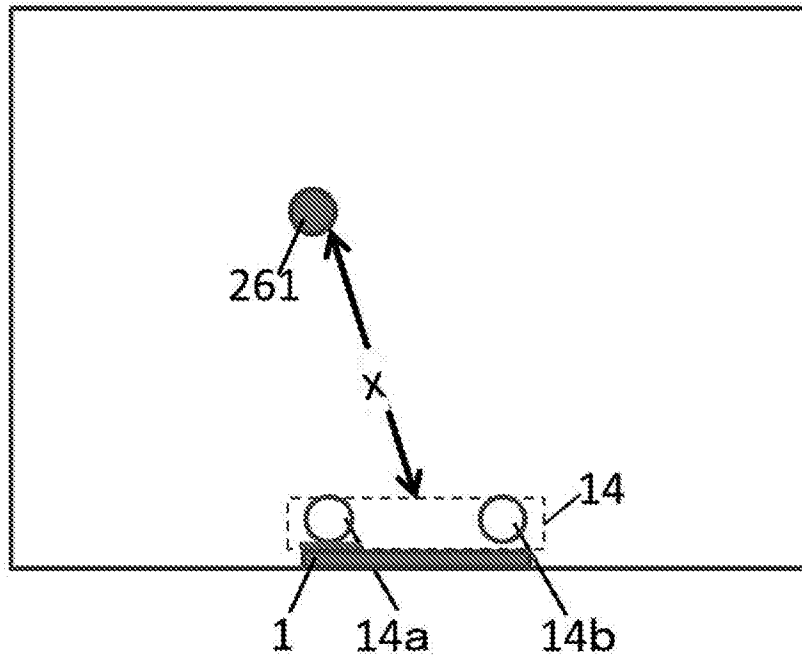
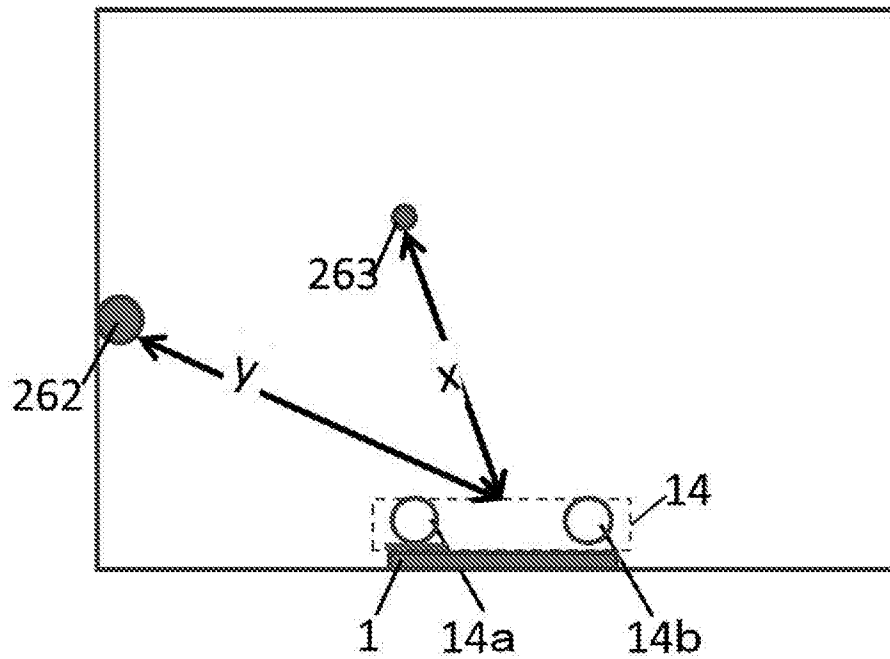


(A)



(B)

[図26]

時刻 t において測定された音源時刻 $t+1$ において測定された音源

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2014/003459

A. CLASSIFICATION OF SUBJECT MATTER

G10L15/28(2013.01)i, G10L25/51(2013.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G10L15/00-15/28, 25/51, H04R1/40, 3/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2014
Kokai Jitsuyo Shinan Koho	1971-2014	Toroku Jitsuyo Shinan Koho	1994-2014

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

IEEE Xplore, CiNii

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	JP 2009-71510 A (National Institute of Advanced Industrial Science and Technology), 02 April 2009 (02.04.2009), entire text; all drawings (Family: none)	1, 2, 5-8, 10-12, 15
Y	WO 2005/048239 A1 (Honda Motor Co., Ltd.), 26 May 2005 (26.05.2005), entire text; all drawings & US 2009/0018828 A1 & EP 1691344 A1 & JP 4516527 B2	1, 2, 5-8, 10-12, 15
Y	JP 2010-217754 A (Toshiba Corp.), 30 September 2010 (30.09.2010), entire text; all drawings & WO 2010/106711 A1 & US 2011/0301950 A1	5, 6, 12

☒ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
22 September, 2014 (22.09.14)

Date of mailing of the international search report
30 September, 2014 (30.09.14)

Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2014/003459

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	JP 2002-34092 A (Sharp Corp.), 31 January 2002 (31.01.2002), entire text; all drawings (Family: none)	6, 10
Y	WO 2012/058229 A1 (QUALCOMM INC.), 03 May 2012 (03.05.2012), entire text; all drawings & JP 2013-545137 A & EP 2633524 A1 & US 2012/0128165 A1 & CN 103189913 A & KR 10-2013-0108391 A	7
Y	JP 2006-58395 A (Kabushiki Kaisha Spectra), 02 March 2006 (02.03.2006), entire text; all drawings (Family: none)	11
A	JP 9-251299 A (Toshiba Corp.), 22 September 1997 (22.09.1997), entire text; all drawings & US 6009396 A & EP 795851 A2	1-15
A	JP 2008-92193 A (Japan Science and Technology Agency), 17 April 2008 (17.04.2008), entire text; all drawings (Family: none)	1-15
A	Yoko SASAKI et al., "Point Sound Source Detection using the Main-Lobe Model of a Microphone Array", Journal of the Robotics Society of Japan, 15 April 2009 (15.04.2009), vol.27, no.3, pages 325 to 333	1-15

A. 発明の属する分野の分類（国際特許分類（IPC））

Int.Cl. G10L15/28(2013.01)i, G10L25/51(2013.01)i

B. 調査を行った分野

調査を行った最小限資料（国際特許分類（IPC））

Int.Cl. G10L15/00-15/28, 25/51, H04R1/40, 3/00

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1922-1996年
日本国公開実用新案公報	1971-2014年
日本国実用新案登録公報	1996-2014年
日本国登録実用新案公報	1994-2014年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

IEEE Xplore, CiNii

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
Y	JP 2009-71510 A（独立行政法人産業技術総合研究所） 2009.04.02, 全文, 全図（ファミリーなし）	1, 2, 5-8, 10-12, 15
Y	WO 2005/048239 A1（本田技研工業株式会社）2005.05.26, 全文, 全 図 & US 2009/0018828 A1 & EP 1691344 A1 & JP 4516527 B2	1, 2, 5-8, 10-12, 15
Y	JP 2010-217754 A（株式会社東芝）2010.09.30, 全文, 全図 & WO 2010/106711 A1 & US 2011/0301950 A1	5, 6, 12

☒ C欄の続きにも文献が列挙されている。☐ パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」特に関連のある文献ではなく、一般的技術水準を示すもの
「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
「O」口頭による開示、使用、展示等に言及する文献
「P」国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
「&」同一パテントファミリー文献

国際調査を完了した日

22.09.2014

国際調査報告の発送日

30.09.2014

国際調査機関の名称及びあて先

日本国特許庁（ISA/J P）
郵便番号100-8915
東京都千代田区霞が関三丁目4番3号

特許庁審査官（権限のある職員）

山下 剛史

5 Z

8946

電話番号 03-3581-1101 内線 3591

C (続き) . 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
Y	JP 2002-34092 A (シャープ株式会社) 2002.01.31, 全文, 全図 (ファミリーなし)	6, 10
Y	WO 2012/058229 A1 (QUALCOMM INCORPORATED) 2012.05.03, 全文, 全図 & JP 2013-545137 A & EP 2633524 A1 & US 2012/0128165 A1 & CN 103189913 A & KR 10-2013-0108391 A	7
Y	JP 2006-58395 A (株式会社スペクトラ) 2006.03.02, 全文, 全図 (ファミリーなし)	11
A	JP 9-251299 A (株式会社東芝) 1997.09.22, 全文, 全図 & US 6009396 A & EP 795851 A2	1-15
A	JP 2008-92193 A (独立行政法人科学技術振興機構) 2008.04.17, 全文, 全図 (ファミリーなし)	1-15
A	佐々木洋子他, マイクアレイのメインローブモデルを用いた点音源 検出手法, 日本ロボット学会誌, 2009.04.15, Vol.27, No.3, p.325-333	1-15