

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7010136号
(P7010136)

(45)発行日 令和4年1月26日(2022.1.26)

(24)登録日 令和4年1月17日(2022.1.17)

(51)国際特許分類		F I			
G 1 0 L	25/51	(2013.01)	G 1 0 L	25/51	4 0 0
G 1 0 L	15/28	(2013.01)	G 1 0 L	15/28	4 0 0

請求項の数 7 (全25頁)

(21)出願番号	特願2018-91943(P2018-91943)	(73)特許権者	000005223 富士通株式会社
(22)出願日	平成30年5月11日(2018.5.11)		神奈川県川崎市中原区上小田中4丁目1番1号
(65)公開番号	特開2019-197179(P2019-197179 A)	(74)代理人	100079049 弁理士 中島 淳
(43)公開日	令和1年11月14日(2019.11.14)	(74)代理人	100084995 弁理士 加藤 和詳
審査請求日	令和3年2月10日(2021.2.10)	(74)代理人	100099025 弁理士 福田 浩志
		(72)発明者	松本 智佳子 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内
		審査官	山下 剛史

最終頁に続く

(54)【発明の名称】 発声方向判定プログラム、発声方向判定方法、及び、発声方向判定装置

(57)【特許請求の範囲】

【請求項1】

1 フレーム毎に、第1音声入力部に入力された第1音声信号に対応する第1周波数信号と第2音声入力部に入力された第2音声信号に対応する第2周波数信号との所定周波数毎の位相差に基づいて発声方向領域を決定し、前記所定周波数毎の位相差が前記発声方向領域から外れている第1位相差外れ度合を算出する処理を複数フレーム分実行し、前記複数フレーム分の前記第1位相差外れ度合から前記複数フレームの統計値である第2位相差外れ度合を算出し、前記第2位相差外れ度合が第1閾値以下である場合、ユーザが前記第1音声入力部及び前記第2音声入力部に向かって発声していると判定する、発声方向判定処理をコンピュータに実行させるためのプログラム。

【請求項2】

前記第1周波数信号は、前記第1音声入力部に入力された1フレーム分の前記第1音声信号を時間周波数変換することで取得され、前記第2周波数信号は、前記第2音声入力部に入力され、前記1フレーム分の第1音声信号に対応する、1フレーム分の第2音声信号を時間周波数変換することで取得される、請求項1に記載のプログラム。

【請求項3】

前記所定周波数毎の位相差に基づいて、前記ユーザが前記第1音声入力部及び前記第2音声入力部の正面に位置する場合の発声方向位相差と、前記ユーザが前記第1音声入力部及

び前記第 2 音声入力部の正面からずれた位置に存在する場合の発声方向位相差と、の差分に基づいて、前記第 1 閾値の値を定める、
請求項 1 または請求項 2 に記載のプログラム。

【請求項 4】

第 2 閾値を超える周波数に対応する前記第 1 周波数信号及び第 2 周波数信号の位相差を使用して前記第 1 位相差外れ度合を算出する、
請求項 1 ~ 請求項 3 の何れか 1 項に記載のプログラム。

【請求項 5】

前記第 1 周波数信号及び前記第 2 周波数信号の少なくとも一方の振幅が第 3 閾値を超える場合に、前記第 1 周波数信号及び第 2 周波数信号の位相差を使用して前記第 1 位相差外れ度合を算出する、
請求項 1 ~ 請求項 4 の何れか 1 項に記載のプログラム。

10

【請求項 6】

コンピュータが、

1 フレーム毎に、第 1 音声入力部に入力された第 1 音声信号に対応する第 1 周波数信号と第 2 音声入力部に入力された第 2 音声信号に対応する第 2 周波数信号との所定周波数毎の位相差に基づいて発声方向領域を決定し、前記所定周波数毎の位相差が前記発声方向領域から外れている第 1 位相差外れ度合を算出する処理を複数フレーム分実行し、
前記複数フレーム分の前記第 1 位相差外れ度合から前記複数フレームの統計値である第 2 位相差外れ度合を算出し、
前記第 2 位相差外れ度合が第 1 閾値以下である場合、ユーザが前記第 1 音声入力部及び前記第 2 音声入力部に向かって発声していると判定する、
発声方向判定方法。

20

【請求項 7】

1 フレーム毎に、第 1 音声入力部に入力された第 1 音声信号に対応する第 1 周波数信号と第 2 音声入力部に入力された第 2 音声信号に対応する第 2 周波数信号との所定周波数毎の位相差に基づいて発声方向領域を決定し、前記所定周波数毎の位相差が前記発声方向領域から外れている第 1 位相差外れ度合を算出する処理を複数フレーム分実行する第 1 位相差外れ算出部と、
前記複数フレーム分の前記第 1 位相差外れ度合から前記複数フレームの統計値である第 2 位相差外れ度合を算出する第 2 位相差外れ算出部と、
前記第 2 位相差外れ度合が第 1 閾値以下である場合、ユーザが前記第 1 音声入力部及び前記第 2 音声入力部に向かって発声していると判定する発声方向判定部と、
を含む、発声方向判定装置。

30

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、発声方向判定プログラム、発声方向判定方法及び発声方向判定装置に関する。

【背景技術】

【0002】

スマートスピーカ等の音声操作装置の普及によって、発声する前に、特別な操作を行うことなく、例えば、ボタンなどを押下することなく、アプリの起動、及び情報検索等を音声で行う機会が増加している。スマートスピーカ等に対しては、音声入力の開始を示すためにボタンなどを押下することなく、発声を行うことが一般的であるため、意図しない発声によって、音声操作装置がユーザの意図しない動作を行う場合がある。

40

【0003】

音声操作装置がユーザの意図しない動作を行わないように、例えば、ユーザがマイクロフォン（以下、マイクという。）に向かって発声しているか否か判定する技術が存在する。当該技術では、例えば、ユーザを撮影した画像データが使用されるが、この場合、画像データを撮影するためにカメラを設置する必要があるため、コストが増大し、システムが複

50

雑化する。また、画像データの撮影には抵抗を感じるユーザも多い。

【先行技術文献】

【特許文献】

【0004】

【文献】特開平10-243494号公報

特開2016-181789号公報

特許5387459号公報

【発明の概要】

【発明が解決しようとする課題】

【0005】

画像データを使用せず、音声データに基づいて、ユーザがマイクに向かって発声しているか否かを判断する技術が存在する。当該技術では、一対のマイクの出力の周波数スペクトルを分析し、10~20[kHz]帯の音声信号のパワー比に基づいて、ユーザがマイクに向かって発声しているか否かを判定する。

【0006】

当該技術では、ユーザが2つのマイクの間位置に存在することを前提としている。したがって、ユーザが移動すると、ユーザの向きを適切に判定することができない。音声の遅延量に基づいてユーザの位置を推定することはできるが、ユーザの移動に伴う補正係数の算出は、ある程度の範囲の位置及び角度を網羅しなければならず、負荷が大きい。

【0007】

本発明は、1つの側面として、ユーザとマイクとの位置関係が変化する場合でも、マイクで取得される音声信号に基づいて、ユーザがマイクに向かって発声しているか否かを判定することを可能とすることを目的とする。

【課題を解決するための手段】

【0008】

1つの実施形態では、1フレーム毎に、第1周波数信号と第2周波数信号との所定周波数毎の位相差に基づいて発声方向領域を決定し、所定周波数毎の位相差が発声方向領域から外れている第1位相差外れ度合を算出する処理を複数フレーム分実行する。第1周波数信号は、第1音声入力部に入力された第1音声信号に対応し、第2周波数信号は、第2音声入力部に入力された第2音声信号に対応する。複数フレーム分の第1位相差外れ度合から複数フレームの統計値である第2位相差外れ度合を算出し、第2位相差外れ度合が第1閾値以下である場合、ユーザが第1音声入力部及び第2音声入力部に向かって発声していると判定する。

【発明の効果】

【0009】

本発明は、1つの側面として、ユーザとマイクとの位置関係が変化する場合でも、マイクで取得される音声信号に基づいて、ユーザがマイクに向かって発声しているか否かを判定することを可能とする。

【図面の簡単な説明】

【0010】

【図1】第1~第4実施形態に係る発声方向判定装置の要部機能の一例を示すブロック図である。

【図2】発声方向領域の決定を説明するための位相差と周波数との関係を例示する概念図である。

【図3A】発声方向領域の決定を説明するための位相差と周波数との関係を例示する概念図である。

【図3B】発声方向領域の決定を説明するための位相差と周波数との関係を例示する概念図である。

【図4A】位相差外れ度合の算出を説明するための位相差と周波数との関係を例示する概念図である。

10

20

30

40

50

【図 4 B】位相差外れ度合及び位相差外れ度合時間平均値と時間との関係を例示する概念図である。

【図 5 A】マイクに対するユーザの顔の向きを例示する概念図である。

【図 5 B】マイクに対するユーザの顔の向きを例示する概念図である。

【図 5 C】マイクに対するユーザの顔の向きを例示する概念図である。

【図 6 A】位相差と周波数との関係を例示する概念図である。

【図 6 B】位相差と周波数との関係を例示する概念図である。

【図 6 C】位相差と周波数との関係を例示する概念図である。

【図 7】位相差外れ度合時間平均値と閾値との関係を例示する概念図である。

【図 8 A】第 1 ~ 第 4 実施形態に係る発声方向判定装置のハードウェアの構成の一例を示すブロック図である。

10

【図 8 B】発声方向判定装置を含む環境を示す概念図である。

【図 9】第 1 実施形態に係る発声方向判定処理の流れの一例を示すフローチャートである。

【図 10】第 1 及び第 2 実施形態に係る位相差外れ度合算出処理の流れの一例を示すフローチャートである。

【図 11 A】マイクに対するユーザの位置及び顔の向きを例示する概念図である。

【図 11 B】マイクに対するユーザの位置及び顔の向きを例示する概念図である。

【図 12 A】閾値の設定を説明するための位相差と周波数との関係を例示する概念図である。

【図 12 B】閾値の設定を説明するための位相差と周波数との関係を例示する概念図である。

20

【図 13】位相差外れ度合時間平均値と閾値との関係を例示する概念図である。

【図 14】閾値と発声方向との関係を例示する概念図である。

【図 15 A】閾値の設定を説明するための位相差と周波数との関係を例示する概念図である。

【図 15 B】閾値の設定を説明するための位相差と周波数との関係を例示する概念図である。

【図 16 A】閾値の設定を説明するための位相差外れ度合時間平均値と時間との関係を例示する概念図である。

【図 16 B】閾値の設定を説明するための位相差外れ度合時間平均値と時間との関係を例示する概念図である。

30

【図 17】第 2 実施形態に係る発声方向判定処理の流れの一例を示すフローチャートである。

【図 18】第 3 実施形態に係る位相差外れ度合算出処理の流れの一例を示すフローチャートである。

【図 19】第 4 実施形態に係る位相差外れ度合算出処理の流れの一例を示すフローチャートである。

【図 20】本実施形態の検証結果を例示する表である。

【図 21 A】本実施形態の検証を説明するための概念図である。

【図 21 B】本実施形態の検証を説明するための概念図である。

40

【発明を実施するための形態】

【0011】

[第 1 実施形態]

以下、図面を参照して第 1 実施形態の一例を詳細に説明する。

【0012】

図 1 に示す発声方向判定装置 10 は、音声入力部 21 A 及び 21 B、時間周波数変換部 22、発声方向領域決定部 24、位相差外れ算出部 25、位相差外れ時間平均値算出部 26、及び発声方向判定部 27 を含む。音声入力部 21 A 及び 21 B は、音声を検出し、検出した音声を音声信号に変換する。以下、音声入力部が 2 個である場合について説明するが、本実施形態はこれに限定されず、音声入力部が 3 個以上含まれていてもよい。

50

【 0 0 1 3 】

時間周波数変換部 2 2 は、音声信号に対して時間周波数変換を行う。時間周波数変換部 2 2 は、例えば、F F T (Fast Fourier Transformation) を使用して、時間に応じて振幅が変化する音声信号 I N T A 及び I N T B を周波数に応じて振幅が変化する周波数信号 I N F A 及び I N F B に変換する。発声方向領域決定部 2 4 は、ユーザが発声している方向であると推定される方向を含む領域である発声方向領域を決定する。

【 0 0 1 4 】

発声方向領域決定部 2 4 は、まず、発声方向を推定する。図 2 は、周波数信号 I N F A 及び I N F B の周波数ビン毎の位相差と周波数との関係を例示する。図 2 の縦軸は、位相差 [r a d] を表わし、横軸は周波数ビン番号を表す。周波数ビン番号の最大値は F F T サイズ / 2 である。F F T サイズは、時間周波数変換に F F T が使用されている場合の 1 フレームあたりのポイント数である。

10

【 0 0 1 5 】

発声方向は、例えば、図 2 に例示するように、周波数ビン番号 0 及び位相差 0 [r a d] の点 (0 , 0) から周波数ビン番号 F F T サイズ / 2 及び位相差 p [r a d] の点 (F F T サイズ / 2 , p) までの直線で表される。位相差 p [r a d] は、周波数ビン番号 F F T サイズ / 2 における発声方向位相差の値である。

【 0 0 1 6 】

発声方向位相差 p [r a d] は、例えば、周波数ビン毎に、位相差 × サンプル周波数 / (2 × 周波数ビン) を算出して加算し、F F T サイズ / 2 で除算することで算出される。図 2 の例では、サンプル周波数 = 2 2 , 0 5 0 [H z] であり、発声方向位相差 p = 1 . 8 6 [r a d] である。

20

【 0 0 1 7 】

発声方向領域決定部 2 4 は、図 3 A に例示するように、直線 7 1 A 及び 7 1 B で囲まれた領域を発声方向領域として決定する。直線 7 1 A は、周波数ビン番号 0 及び位相差 0 [r a d] の点 (0 , 0) から周波数ビン番号 F F T サイズ / 2 及び位相差 p + の点 (F F T サイズ / 2 , p +) までの直線である。直線 7 1 B は、周波数ビン番号 0 及び位相差 0 [r a d] の点 (0 , 0) から周波数ビン番号 F F T サイズ / 2 及び位相差 p - の点 (F F T サイズ / 2 , p -) までの直線である。は、例えば、1 . 0 [r a d] であってよい。

30

【 0 0 1 8 】

図 3 A は、発声方向位相差 p が 0 [r a d] である例であり、図 3 B は、発声方向位相差 p が 1 . 8 6 [r a d] である例である。なお、例えば、予め分割された位相差の領域を複数用意し、位相差を用いて各領域に音源が存在する率を算出し、領域を選別することで、発声方向及び発声方向領域を同時に決定するようにしてもよい。

【 0 0 1 9 】

位相差外れ算出部 2 5 は、第 1 位相差外れ算出部の一例であり、フレーム毎の位相差外れ度合を算出する。位相差外れ度合は第 1 位相差外れ度合の一例である。1 フレームの長さは、例えば、約 1 0 [m s] であってよい。フレーム毎の位相差外れ度合 D f n は、例えば、以下の (1) 式で算出される。なお、以下で、フレーム番号を限定しない場合、位相差外れ度合を D で表す。

40

【 数 1 】

$$D_{fn} = \frac{\sum_{fr=0}^{FR-1} \Delta d_{fr}}{FR} \quad \dots (1)$$

【 0 0 2 0 】

F R は、最大周波数ビン番号であり、時間周波数変換が F F T で行われる場合、F F T サイズ / 2 である。個別位相差外れ度合 d f r は、図 4 A に例示するように、周波数ビン

50

番号 f_r で、周波数信号 $I N F A$ 及び $I N F B$ の位相差が発声方向領域 $81R$ から外れている度合、即ち、発声方向領域 $81R$ の端部と位相差との縦軸方向の距離である。なお、図 4 A は $p = 0$ 、 $q = 1$ の例であり、発声方向領域 $81R$ は点 $(0, 0)$ 、点 $(FFT サイズ / 2, 1)$ 、点 $(FFT サイズ / 2, -1)$ で囲まれた領域である。

【0021】

なお、以下で、周波数ビン番号を限定しない場合、個別位相差外れ度合を d で表す。周波数信号 $I N F A$ 及び $I N F B$ の位相差が発声方向領域 $81R$ 内に存在する場合、 d を 0 とする。

【0022】

位相差外れ算出部 25 は、個別位相差外れ度合 d_{f_r} を周波数ビン f_r 毎に算出し、1 フレーム内の個別位相差外れ度合 d の平均値である位相差外れ度合 D_{f_n} をフレーム毎に算出する。 f_n はフレーム番号を表す。なお、位相差外れ度合 D_{f_n} は、単に、1 フレーム内の個別位相差外れ度合 d を合計した値でもよい。

10

【0023】

位相差外れ時間平均値算出部 26 は、第 2 位相差外れ算出部の一例であり、位相差外れ度合 D の所定数フレーム毎の平均値である位相差外れ度合時間平均値 $A D_{f_n}$ を算出する。位相差外れ度合時間平均値 $A D_{f_n}$ は、第 2 位相差外れ度合の一例であり、以下の (2) 式で算出される。 f_c は、所定数フレームの個数であり、例えば、10 であってよい。 $A D_{f_n}$ は、フレーム番号 f_n から過去 f_c フレーム分の位相差外れ度合 D の平均値である。なお、以下で、フレーム番号を限定しない場合、位相差外れ度合時間平均値を $A D$ で表す。

20

【数 2】

$$AD_{f_n} = \frac{\sum_{i=0}^{f_c-1} D_{f_n-i}}{f_c} \quad \dots (2)$$

【0024】

図 4 B は、位相差外れ度合 D と時間との関係及び位相差外れ度合時間平均値 $A D$ と時間との関係を例示する。図 4 B の縦軸は、位相差外れ度合 D または位相差外れ度合時間平均値 $A D$ を表し、横軸は時間、即ち、フレーム番号 f_n を表す。図 4 B において、位相差外れ度合 D は棒 82D で表され、位相差外れ度合時間平均値 $A D$ は線 82AD で表される。

30

【0025】

発声方向判定部 27 は、ユーザが音声入力部 21A 及び 21B に向かって発声しているか否か、を判定する。即ち、発声方向判定部 27 は、発声しているユーザの顔、詳細には、ユーザの口がマイクロフォン（以下、マイクという。）55A 及び 55B に向けられているか、発声しているユーザの顔がマイク 55A 及び 55B に向けられていないか、を判定する。マイク 55A 及び 55B は、音声入力部 21A 及び 21B の一例である。以下、マイク 55A 及び 55B を区別しない場合、マイク 55 という。図 5 A は、ユーザの顔がマイク 55 に向けられている例であり、図 5 B 及び図 5 C は、ユーザの顔がマイク 55 に向けられていない例である。

40

【0026】

図 5 A に例示するように、マイク 55A 及びマイク 55B は、発声方向判定装置 10 に所定の間隔を空けて配置される。ユーザの顔がマイク 55 に向かっている、とは、詳細には、ユーザの顔がマイク 55A 及び 55B の中間位置を中心とする所定範囲の領域に向けられている、ことを示す。

【0027】

発声方向判定部 27 は、位相差外れ度合時間平均値 $A D$ が所定閾値以下である場合、ユーザがマイク 55 に向かって発声していると判定する。所定閾値は、第 1 閾値の一例であり、例えば、 $0.6 [rad]$ であってよいが、本実施形態はこれに限定されない。

50

【 0 0 2 8 】

図 6 A は、図 5 A に例示するように、ユーザがマイク 5 5 に向かって発声している場合の周波数信号 I N F A 及び I N F B の周波数ビン毎の位相差と周波数との関係を例示する。但し、図 6 A では、特徴を顕著に表すために、複数フレームの周波数信号 I N F A 及び I N F B の位相差を重畳している。また、図 6 A、図 6 B、及び図 6 C の発声方向領域 8 3 は、図 4 A の発声方向領域 8 1 R と同様の例である。

【 0 0 2 9 】

図 6 B は、図 5 B に例示するように、ユーザがマイク 5 5 に向かわず、マイク 5 5 の左側に向かって発声している場合の位相差と周波数との関係を例示する。図 6 C は、図 5 C に例示するように、ユーザがマイク 5 5 に向かわず、ユーザの右手方向に向かって発声している場合の周波数信号 I N F A 及び I N F B の周波数ビン毎の位相差と周波数との関係を例示する。

10

【 0 0 3 0 】

図 6 A と比較して、図 6 B 及び図 6 C では、発声方向領域 8 3 から外れている、周波数信号 I N F A 及び I N F B の位相差が多い。即ち、図 5 B 及び図 5 C に例示するように、ユーザがマイク 5 5 に向かわずに発声している場合、図 5 A に例示するように、ユーザがマイク 5 5 に向かって発声している場合よりも、位相差外れ度合時間平均値 A D は大きい。

【 0 0 3 1 】

図 7 に、図 6 A、図 6 B 及び図 6 C に各々対応する位相差外れ度合時間平均値 8 4 C、8 4 L 及び 8 4 R、を例示する。図 7 に例示されるように、図 6 A に対応する位相差外れ度合時間平均値 8 4 C と、図 6 B 及び図 6 C に対応する位相差外れ度合時間平均値 8 4 L 及び 8 4 R と、は全体的に大きさが異なる。即ち、位相差外れ度合時間平均値が所定閾値 8 4 T 以下であるか否かが判定することで、ユーザの顔の向きを判定することができる。

20

【 0 0 3 2 】

マイクに対するユーザの位置が変化しない場合、ユーザがマイクに向かって発声すると反射音が直接音より少なく、ユーザがマイクに向かわず発声すると反射音が直接音より多い。即ち、ユーザがマイクに向かわず発声すると、マイクに検出される反射音が多くなるため、比較的大きい個別位相差外れ度合が多く出現し、位相差外れ度合時間平均値が大きくなる。したがって、位相差外れ度合時間平均値が所定閾値以下である場合、ユーザがマイクに向かって発声していると判定することができる。

30

【 0 0 3 3 】

発声方向判定装置 1 0 は、一例として、図 8 A に示すように、C P U (Central Processing Unit) 5 1、一次記憶部 5 2、二次記憶部 5 3、及び、外部インターフェイス 5 4 を含む。C P U 5 1 は、ハードウェアであるプロセッサの一例である。C P U 5 1、一次記憶部 5 2、二次記憶部 5 3、及び、外部インターフェイス 5 4 は、バス 5 9 を介して相互に接続されている。

【 0 0 3 4 】

一次記憶部 5 2 は、例えば、R A M (Random Access Memory) などの揮発性のメモリである。二次記憶部 5 3 は、例えば、H D D (Hard Disk Drive)、又は S S D (Solid State Drive) などの不揮発性のメモリである。

40

【 0 0 3 5 】

二次記憶部 5 3 は、プログラム格納領域 5 3 A 及びデータ格納領域 5 3 B を含む。プログラム格納領域 5 3 A は、一例として、発声方向判定プログラムなどのプログラムを記憶している。データ格納領域 5 3 B は、一例として、音声信号及び発声方向判定プログラムを実行している間に生成される中間データなどを記憶する。

【 0 0 3 6 】

C P U 5 1 は、プログラム格納領域 5 3 A から発声方向判定プログラムを読み出して一次記憶部 5 2 に展開する。C P U 5 1 は、発声方向判定プログラムを実行することで、図 1 の時間周波数変換部 2 2、発声方向領域決定部 2 4、位相差外れ算出部 2 5、位相差外れ時間平均値算出部 2 6、及び発声方向判定部 2 7 として動作する。

50

【 0 0 3 7 】

なお、発声方向判定プログラムなどのプログラムは、外部サーバに記憶され、ネットワークを介して、一次記憶部 5 2 に展開されてもよい。また、発声方向判定プログラムなどのプログラムは、DVD (Digital Versatile Disc) などの非一時的記録媒体に記憶され、記録媒体読込装置を介して、一次記憶部 5 2 に展開されてもよい。

【 0 0 3 8 】

外部インターフェイス 5 4 には外部装置が接続され、外部インターフェイス 5 4 は、外部装置と CPU 5 1 との間の各種情報の送受信を司る。外部インターフェイス 5 4 には、マイク 5 5 A 及び 5 5 B、及びスピーカ 5 6 が接続されている。マイク 5 5 A 及び 5 5 B、及びスピーカ 5 6 は、外部装置であることに限定されず、発声方向判定装置 1 0 に内蔵されていてもよい。

10

【 0 0 3 9 】

マイク 5 5 A 及び 5 5 B は、音声入力部 2 1 A 及び 2 1 B の一例であり、ユーザが発話した音声を検出し、音声信号に変換する。スピーカ 5 6 は、例えば、発声方向判定装置 1 0 が出力する発声方向判定結果を音声で報知する。

【 0 0 4 0 】

発声方向判定装置 1 0 は、例えば、スマートスピーカなどの音声操作装置に内蔵されていてもよいし、図 8 B に例示するように、音声操作装置 6 1 に近接して配置されていてもよい。図 8 B は、音声操作装置 6 1 の上面に発声方向判定装置 1 0 を配置した例であるが、例えば、音声操作装置 6 1 の前面に発声方向判定装置 1 0 を配置してもよい。音声操作装置 6 1 に近接して配置されている場合、発声方向判定装置 1 0 と音声操作装置 6 1 とは有線または無線で接続される。

20

【 0 0 4 1 】

次に、発声方向判定処理の作用の概要について説明する。図 9 は、発声方向判定処理の流れを例示する。CPU 5 1 は、ステップ 1 0 1 で、マイク 5 5 A 及びマイク 5 5 B で検出された音声に対応する音声信号 INT A 及び INT B を各々 1 フレーム分読み込む。

【 0 0 4 2 】

CPU 5 1 は、ステップ 1 0 2 で、読み込まれた音声信号 INT A 及び INT B を、周波数信号 INF A 及び INF B に各々時間周波数変換する。ここでは、時間周波数変換が、FFT で行われる例について説明する。CPU 5 1 は、ステップ 1 0 3 で、周波数信号 INF A 及び INF B の位相差を算出する。

30

【 0 0 4 3 】

CPU 5 1 は、ステップ 1 0 4 で、発声方向領域を決定する。CPU 5 1 は、ステップ 1 0 5 で、後述するように、位相差外れ度合平均値を算出する。CPU 5 1 は、ステップ 1 0 6 で、ステップ 1 0 1 で読み込んだフレームの数が所定フレーム数 M を超えたか否か判定する。ステップ 1 0 6 の判定が否定された場合、即ち、ステップ 1 0 1 で読み込んだフレームの数が所定フレーム数 M を超えていない場合、CPU 5 1 は、ステップ 1 0 1 に戻る。

【 0 0 4 4 】

ステップ 1 0 6 の判定が肯定された場合、即ち、ステップ 1 0 1 で読み込んだフレームの数が所定フレーム数 M を超えている場合、CPU 5 1 は、ステップ 1 0 7 で、位相差外れ度合時間平均値を算出する。CPU 5 1 は、ステップ 1 0 8 で、位相差外れ度合時間平均値が所定閾値以下であるか否か判定する。

40

【 0 0 4 5 】

ステップ 1 0 8 の判定が肯定された場合、即ち、位相差外れ度合時間平均値が所定閾値以下である場合、CPU 5 1 は、ステップ 1 0 9 で、ユーザの顔がマイクに向けられていると判定し、例えば、発声方向判定フラグに「1」をセットする。ステップ 1 0 8 の判定が否定された場合、即ち、位相差外れ度合時間平均値が所定閾値を超えている場合、CPU 5 1 は、ステップ 1 1 0 で、ユーザの顔がマイクに向けられていないと判定し、例えば、発声方向判定フラグに「0」をセットする。発声方向判定フラグは、図 8 A の二次記憶部

50

5 3 のデータ格納領域 5 3 B に含まれていてもよい

【 0 0 4 6 】

C P U 5 1 は、ステップ 1 1 1 で、例えば、所定のボタンが押下されるなど、ユーザによって発声方向判定処理の終了指示が行われたか否か判定する。ステップ 1 1 1 の判定が否定された場合、C P U 5 1 はステップ 1 0 1 に戻る。

【 0 0 4 7 】

ステップ 1 1 1 の判定が肯定された場合、C P U 5 1 は、発声方向判定フラグの値に基づいて、判定結果を示す信号を出力して発声方向判定処理を終了する。判定結果を示す信号は、例えば、スマートスピーカなどの音声操作装置 6 1 に入力される。音声操作装置 6 1 は、例えば、判定結果が、ユーザの顔がマイクに向けられていることを示す場合、ユーザの発声にしたがって作動し、判定結果が、ユーザの顔がマイクに向けられていないことを示す場合、ユーザの発声を無視する。

10

【 0 0 4 8 】

即ち、ユーザがマイクに顔を向けて発声していると判定された場合に音声を受け付け、ユーザがマイクに顔を向けていないと判定された場合に音声を受け付けないことが可能となるため、ユーザが意図しない誤認識、誤動作を低減することが可能となる。

【 0 0 4 9 】

判定結果を示す信号は、例えば、テレビ会議システムに入力されてもよい。テレビ会議システムは、例えば、判定結果が、ユーザの顔がマイクに向けられていることを示す場合、ユーザの発言を他の出席者に送信し、判定結果が、ユーザの顔がマイクに向けられていないことを示す場合、ユーザの発声を無視する。

20

【 0 0 5 0 】

判定結果を示す信号は、例えば、音声認識装置に入力されてもよい。音声認識装置は、例えば、判定結果が、ユーザの顔がマイクに向けられていることを示す場合、ユーザの発声の音声認識を行い、判定結果が、ユーザの顔がマイクに向けられていないことを示す場合、ユーザの発声を無視する。

【 0 0 5 1 】

判定結果を示す信号は、例えば、安全確認システムに入力されてもよい。安全確認システムは、例えば、判定結果が、ユーザの顔がマイクに向けられていることを示す場合、ユーザが安全であると判定する。安全確認システムは、例えば、判定結果が、所定時間の間、ユーザの顔がマイクに向けられていないことを示す場合、ユーザに対して、マイクに向かって発声するように、スピーカ 5 6、または、文字または画像を出力する画像出力装置を介して報知してもよい。

30

【 0 0 5 2 】

図 1 0 は、図 9 のステップ 1 0 5 の位相差外れ度合算出処理の流れを例示する。C P U 5 1 は、ステップ 2 0 1 で、個別位相差外れ度合合計値を記憶する d 合計及び周波数ピン数をカウントする変数 i に 0 をセットする。C P U 5 1 は、ステップ 2 0 2 で、周波数信号 I N F A 及び I N F B の位相差が発声方向領域から外れている度合である個別位相差外れ度合 d を周波数ピン番号 i 毎に算出する。周波数信号 I N F A 及び I N F B の位相差が発声方向領域内に存在する場合、 $d = 0$ とする。

40

【 0 0 5 3 】

C P U 5 1 は、ステップ 2 0 3 で、 d 合計の値に d の値を加算し、変数 i の値に 1 を加算する。C P U 5 1 は、ステップ 2 0 4 で、変数 i の値が最大周波数ピン番号、即ち、F F T サイズ / 2 未満であるか否か判定する。ステップ 2 0 4 の判定が肯定された場合、即ち、変数 i の値が F F T サイズ / 2 未満である場合、C P U 5 1 は、ステップ 2 0 2 に戻る。ステップ 2 0 4 の判定が否定された場合、即ち、変数 i の値が F F T サイズ / 2 以上である場合、C P U 5 1 は、ステップ 2 0 5 で、 d 合計の値を変数 i の値で除算し、位相差外れ度合平均値を算出する。

【 0 0 5 4 】

なお、ステップ 1 0 8 では、M フレーム分の位相差外れ度合から算出される位相差外れ度

50

合時間平均値に基づいて、ユーザがマイクに向かって発声しているか否か判定している。しかしながら、ステップ108の判定を複数回繰り返し、位相差外れ度合時間平均値が所定閾値以下であると判定された回数が所定数を超える場合に、ユーザがマイクに向かって発声していると判定してもよい。また、ステップ108の判定を複数回繰り返し、位相差外れ度合時間平均値が所定閾値以下であると判定された回数が所定の割合を超える場合に、ユーザがマイクに向かって発声していると判定してもよい。また、Mフレーム分の位相差外れ度合から算出される位相差外れ度合時間平均値に代えて、例えば、Mフレーム分の位相差外れ度合から算出される個別位相差外れ度合中央値などの個別位相差外れ度合の統計値を使用してもよい。

【0055】

本実施形態では、1フレーム毎に、第1周波数信号と第2周波数信号との所定周波数毎の位相差に基づいて発声方向領域を決定し、所定周波数毎の位相差が発声方向領域から外れている第1位相差外れ度合を算出する処理を複数フレーム分実行する。第1周波数信号は、第1音声入力部に入力された第1音声信号に対応し、第2周波数信号は、第2音声入力部に入力された第2音声信号に対応する。複数フレーム分の第1位相差外れ度合から複数フレームの統計値である第2位相差外れ度合を算出し、第2位相差外れ度合が第1閾値以下である場合、ユーザが第1音声入力部及び第2音声入力部に向かって発声していると判定する。

【0056】

これにより、ユーザとマイクとの位置関係が変化した場合でも、マイクで検出される音声信号に基づいて、ユーザがマイクに向かって発声しているか否かを判定することができる。

【0057】

[第2実施形態]

次に、第2実施形態の一例を説明する。第1実施形態と同様の構成及び作用については、説明を省略する。第2実施形態では、推定される発声方向に基づいて、閾値を設定する点で、第1実施形態と異なる。

【0058】

図11Aに例示するように、ユーザがマイク55の正面からユーザの右手方向にずれて位置し、マイク55に顔を向けて発声する場合の位相差と周波数との関係を、図12Aに例示する。図11Bに例示するように、ユーザがマイク55の正面からユーザの右手方向にずれて位置し、さらに、顔を右側に向けて、即ち、マイク55に向かわないで発声する場合の位相差と周波数との関係を、図12Bに例示する。

【0059】

ユーザがマイク55の正面に位置する場合の図6A及び図6Cと比較して、図12A及び図12Bでは、発声方向領域から外れている周波数信号INF A及びINF Bの位相差が多く、個別位相差外れ度合も大きい。

【0060】

図13に、位相差外れ度合時間平均値と時間との関係を表すグラフを例示する。線85Rは、図11Bに例示するように、ユーザの顔がマイク55に向けられていない場合の位相差外れ度合時間平均値を表す。また、線85Cは、図11Aに例示するように、ユーザの顔がマイク55に向けられている場合の位相差外れ度合時間平均値を表す。

【0061】

線85Rは、図7の線84Rよりも全体的に大きい値を表し、線85Cは、図7の線84Cよりも全体的に大きい値を表す。したがって、図7の閾値84Tより大きい閾値を使用することで、線85Rと線85Cとを適切に区別する、即ち、ユーザの顔の向きを判定する、ことが可能となる。

【0062】

図14に、閾値とユーザの位置との関係を表すグラフを例示する。図14の縦軸は閾値[r a d]を表し、横軸はユーザの位置である発声方向を表す。ユーザがマイク55の正面に位置する場合、発声方向は0で表される。発声方向は、ユーザがマイク55の正面から

10

20

30

40

50

ユーザの右手方向または左手方向に向かって遠ざかる程大きい値で表される。即ち、閾値は、ユーザの位置がマイクの正面から右手方向または左手方向に向かって遠ざかる程、大きい値に設定される。

【 0 0 6 3 】

例えば、閾値は (3) 式で算出される。

閾値 = 調整値 × p + 基本閾値 ... (3)

基本閾値は、図 1 4 において、ユーザがマイク 5 5 の正面に位置する場合、即ち、発声方向が 0 の場合の閾値であり、調整値は、図 1 4 の閾値を表す線の傾きである。

【 0 0 6 4 】

例えば、調整値が 0 . 2 5、基本閾値が 0 . 6 [r a d] であり、図 1 5 A に例示するように、発声方向位相差 p = 0 . 0 である場合、即ち、ユーザがマイク 5 5 の正面に位置する場合、閾値は 0 . 6 (= 0 . 2 5 × 0 . 0 + 0 . 6) [r a d] である。即ち、図 1 6 A に例示するように、閾値 8 6 T は 0 . 6 に設定され、線 8 6 R 及び線 8 6 L と、線 8 6 C で例示するユーザが顔をマイクに向けている場合の位相差外れ度合時間平均値と、を区別する。線 8 6 R は、ユーザが顔を右に向けている場合の位相差外れ度合時間平均値を例示し、線 8 6 L は、ユーザが顔を左に向けている場合の位相差外れ度合時間平均値を例示する。

10

【 0 0 6 5 】

例えば、調整値が 0 . 2 5、基本閾値が 0 . 6 [r a d] であり、図 1 5 B に例示するように、発声方向位相差 p = 1 . 6 である場合、即ち、ユーザがマイク 5 5 の正面からユーザの右手方向にずれて位置する場合、閾値は 1 . 0 (= 0 . 2 5 × 1 . 6 + 0 . 6) である。図 1 6 B に例示するように、閾値 6 4 T は、閾値 1 . 0 に設定され、線 6 4 R で例示するユーザが顔を右に向けている場合の位相差外れ度合時間平均値と、線 6 4 C で例示するユーザが顔をマイクに向けている場合の位相差外れ度合時間平均値と、を区別する。

20

【 0 0 6 6 】

図 1 7 は、第 2 実施形態の発声方向判定処理の流れを例示する。ステップ 1 2 1 ~ ステップ 1 2 7 は、図 9 のステップ 1 0 1 ~ 1 0 7 と同様であるため、説明を省略する。CPU 5 1 は、ステップ 1 2 8 で、例えば、(3) 式を使用して、閾値を設定する。ステップ 1 2 9 は、ステップ 1 2 8 で設定した閾値を使用して、位相差外れ度合時間平均値が閾値を超えるか否かが判定する。ステップ 1 3 0 ~ ステップ 1 3 3 は、図 9 のステップ 1 0 9 ~ ステップ 1 1 2 と同様であるため、説明を省略する。

30

【 0 0 6 7 】

第 2 実施形態では、周波数ビン毎の位相差に基づいて、ユーザがマイクの正面に位置する場合の発声方向位相差と、ユーザがマイクの正面からずれた位置に存在する場合の発声方向位相差と、の差分に基づいて、閾値を設定する。

【 0 0 6 8 】

これにより、ユーザがマイクの正面からずれて位置することによる、位相差外れ度合時間平均値の増大を吸収することができる。したがって、ユーザとマイクとの位置関係が変化した場合でも、マイクで取得された音声信号に基づいて、ユーザがマイクに向かって発声しているか否かを適切に判定することができる。

40

【 0 0 6 9 】

[第 3 実施形態]

次に、第 3 実施形態の一例を説明する。第 1 実施形態及び第 2 実施形態と同様の構成及び作用については、説明を省略する。第 3 実施形態では、所定周波数以上の周波数信号の位相差を使用して位相差外れ度合を算出する点で、第 1 実施形態及び第 2 実施形態と異なる。

【 0 0 7 0 】

第 3 実施形態では、低域周波数帯と高域周波数帯とを区別する閾値 F t h を設定する。閾値 F t h は、第 2 閾値の一例であり、一定の値、例えば、1 0 0 [k H z] であってよい。しかしながら、例えば、ユーザが発声している間、仮閾値 P F t h を自動的に変動させ、位相差外れ度合時間平均値を算出し、位相差外れ度合時間平均値が所定時間の間 0 . 5

50

[r a d] 以下になる仮閾値 P F t h を閾値 F t h として設定してもよい。

【 0 0 7 1 】

図 1 8 は、図 9 のステップ 1 0 5 または図 1 7 のステップ 1 2 5 の位相差外れ度合算出処理の流れを例示する。C P U 5 1 は、ステップ 2 1 1 で、個別位相差外れ度合 d の合計を算出するための d 合計、周波数ピン数をカウントする変数 i、及び高域周波数の周波数ピン数をカウントする変数 j に 0 をセットする。C P U 5 1 は、ステップ 2 1 2 で、変数 i の値に対応する周波数ピンに対応する周波数が所定周波数 F t h 以上であるか否かを判定する。

【 0 0 7 2 】

ステップ 2 1 2 の判定が否定された場合、即ち、変数 i の値に対応する周波数ピンに対応する周波数が所定周波数 F t h 未満である場合、C P U 5 1 は、ステップ 2 1 5 に進む。ステップ 2 1 2 の判定が肯定された場合、即ち、変数 i の値に対応する周波数ピンの周波数が所定周波数 F t h 以上である場合、C P U 5 1 は、個別位相差外れ度合 d を算出する。個別位相差外れ度合 d は、周波数ピン番号 i の周波数信号 I N F A 及び I N F B の位相差が発声方向領域から外れている度合である。

10

【 0 0 7 3 】

C P U 5 1 は、ステップ 2 1 4 で、d 合計の値に d を加算し、変数 j の値に 1 を加算する。C P U 5 1 は、ステップ 2 1 5 で、変数 i の値に 1 を加算する。C P U 5 1 は、ステップ 2 1 6 で、変数 i の値が最大周波数ピン番号、即ち、F F T サイズ / 2 未満であるか否かを判定する。ステップ 2 1 6 の判定が肯定された場合、C P U 5 1 は、ステップ 2 1 2 に戻る。ステップ 2 1 6 の判定が否定された場合、C P U 5 1 は、ステップ 2 1 7 で、d 合計の値を変数 j の値で除算した値を位相差外れ度合として算出する。

20

【 0 0 7 4 】

第 3 実施形態は、第 1 実施形態または第 2 実施形態と組み合わせられてもよい。

【 0 0 7 5 】

ユーザとマイクとの位置関係が変化しない場合、ユーザがマイクに向かって発声すると反射音が直接音より少なく、ユーザがマイクに向かわず発声すると反射音が直接音より多い。この特徴は、中高域の周波数帯において顕著である。また、低域周波数帯の周波数信号は、理論値と実測値との差が大きい場合がある。

【 0 0 7 6 】

したがって、第 3 実施形態では、低域周波数帯の周波数信号を使用せず、所定周波数以上の高域周波数帯の周波数信号の位相差を使用して、位相差外れ度合を算出することで、より精度の高い位相差外れ度合時間平均値を算出することができる。これにより、ユーザとマイクとの位置関係が変化した場合でも、マイクで取得される音声信号に基づいて、ユーザがマイクに向かって発声しているか否かをより適切に判定することができる。

30

【 0 0 7 7 】

[第 4 実施形態]

次に、第 4 実施形態の一例を説明する。第 1 ~ 第 3 実施形態と同様の構成及び作用については、説明を省略する。第 4 実施形態では、周波数信号の振幅が所定の振幅閾値を超える周波数信号の位相差を使用して位相差外れ度合を算出する点で、第 1 ~ 第 3 実施形態と異なる。

40

【 0 0 7 8 】

第 4 実施形態では、例えば、マイク 5 5 A で検出された音声信号 I N T A に対応する周波数信号 I N F A の振幅が振幅閾値 V t h を超える場合、周波数信号 I N F A 及び I N F B の位相差を使用して、位相差外れ度合を算出する。しかしながら、マイク 5 5 B で検出された音声信号 I N T B に対応する周波数信号 I N F B の振幅が振幅閾値 V t h を超える場合に、周波数信号 I N F A 及び I N F B の位相差を使用して、位相差外れ度合を算出するようにしてもよい。また、周波数信号 I N F A 及び I N F B の振幅が両方とも振幅閾値 V t h を超える場合に、周波数信号 I N F A 及び I N F B の位相差を使用して、位相差外れ度合を算出するようにしてもよい。振幅閾値 V t h は、第 3 閾値の一例であり、例えば、

50

定常雑音レベルであってよい。定常雑音レベルは、既存の方法で算出することができる。

【0079】

図19は、図9のステップ105または図17のステップ125の位相差外れ度合算出処理の流れを例示する。CPU51は、ステップ211で、個別位相差外れ度合 d の合計を算出するための d 合計、周波数ピン数をカウントする変数 i 及び振幅閾値 V_{th} を超える周波数信号の周波数ピン数をカウントする変数 k に0をセットする。

【0080】

CPU51は、ステップ222で、例えば、マイク55Aで検出された音声信号に対応する周波数信号の振幅が振幅閾値 V_{th} を超えているか否か判定する。ステップ222の判定が否定された場合、CPU51は、ステップ225に進む。ステップ222の判定が肯定された場合、CPU51は、発声方向領域から周波数ピン番号 i の周波数信号 $INFA$ 及び $INFB$ の位相差が外れている度合である個別位相差外れ度合 d を算出する。

10

【0081】

CPU51は、ステップ224で、 d 合計の値に d を加算し、変数 k の値に1を加算する。CPU51は、ステップ225で、変数 i の値に1を加算する。CPU51は、ステップ226で、変数 i の値が最大周波数ピン番号、即ち、FFTサイズ/2未満であるか否か判定する。ステップ226の判定が肯定された場合、即ち、変数 i の値が最大周波数ピン番号未満である場合、CPU51は、ステップ222に戻る。ステップ226の判定が否定された場合、変数 i の値が最大周波数ピン番号以上である場合、CPU51は、ステップ227で、 d 合計の値を k の値で除算した値を位相差外れ度合として算出する。

20

【0082】

なお、本実施形態では、周波数信号の振幅が振幅閾値を超える周波数信号の位相差を使用して、位相差外れ度合を算出する例について説明したが、本実施形態はこれに限定されない。例えば、音声信号の振幅が所定閾値を超える場合に、音声信号に対応する周波数信号の位相差を使用して、位相差外れ度合を算出するようにしてもよい。また、周波数信号または音声信号の振幅に代えて、パワーが所定閾値を超える周波数信号の位相差を使用して、位相差外れ度合を算出してもよい。

【0083】

第4実施形態は、第1実施形態、第2実施形態、第1実施形態と第3実施形態との組み合わせ、または、第2実施形態と第3実施形態との組み合わせに適用されてもよい。

30

【0084】

第4実施形態では、周波数信号の振幅が所定の振幅閾値を超える場合、位相差外れ度合を算出することで、雑音ではない可能性が高い周波数信号に基づいて、より精度の高い位相差外れ度合時間平均値を算出することができる。これにより、ユーザとマイクとの位置関係が変化した場合でも、マイクで取得された音声信号に基づいて、ユーザがマイクに向かって発声しているか否かをより適切に判定することができる。

【0085】

図9、10、17～19のフローチャートは一例であり、処理の順序は適宜変更可能である。

【0086】

[検証例]

本実施形態による発声方向判定の判定結果を図20に例示する。この検証では、ユーザがマイクから50cm離れた位置から40秒間発声し、サンプリング周波数は22[kHz]である。

40

【0087】

図21Aに例示するユーザ91A及び図21Bに例示するユーザ91Cは、マイク55の正面、即ち、マイク55A及び55Bの中間位置から発声方向判定装置10の正面に向かって延びる直線上に存在する。ユーザ91B及び91Dは、マイク55の正面からずれた位置、即ち、マイク55A及び55Bの中間位置から発声方向判定装置10の正面に向かって延びる直線と、マイク55A及び55Bの中心を交点として、60°の角度で交差す

50

る直線上に存在する。

【 0 0 8 8 】

ユーザ 9 1 A のように、ユーザがマイク 5 5 の正面に位置し、マイク 5 5 に顔を向けている場合、ユーザの顔の向きが適切に判定される割合は 9 5 % である。ユーザ 9 1 B のように、マイク 5 5 の正面からユーザの右手方向にずれて位置し、ユーザがマイク 5 5 に顔を向けている場合、ユーザの顔の向きが適切に判定される割合は 9 3 % である。

【 0 0 8 9 】

ユーザ 9 1 C のように、マイク 5 5 の正面に位置し、ユーザの左手方向に顔を向けている場合、即ち、ユーザがマイク 5 5 に顔を向けていない場合、ユーザの顔の向きが適切に判定される割合は 9 2 % である。ユーザ 9 1 D のように、マイク 5 5 の正面からユーザの右手方向にずれて位置し、ユーザの右手方向に顔を向けている場合、即ち、ユーザがマイク 5 5 に顔を向けていない場合、ユーザの顔の向きが適切に判定される割合は、9 1 % である。

10

【 0 0 9 0 】

即ち、本実施形態によれば、何れの場合でも、9 0 % 以上の割合で、ユーザの位置及び顔の向きが適切に判定される。

【 0 0 9 1 】

以上の各実施形態に関し、更に以下の付記を開示する。

【 0 0 9 2 】

(付記 1)

1 フレーム毎に、第 1 音声入力部に入力された第 1 音声信号に対応する第 1 周波数信号と第 2 音声入力部に入力された第 2 音声信号に対応する第 2 周波数信号との所定周波数毎の位相差に基づいて発声方向領域を決定し、前記所定周波数毎の位相差が前記発声方向領域から外れている第 1 位相差外れ度合を算出する処理を複数フレーム分実行し、前記複数フレーム分の前記第 1 位相差外れ度合から前記複数フレームの統計値である第 2 位相差外れ度合を算出し、前記第 2 位相差外れ度合が第 1 閾値以下である場合、ユーザが前記第 1 音声入力部及び前記第 2 音声入力部に向かって発声していると判定する、発声方向判定処理をコンピュータに実行させるためのプログラム。

20

(付記 2)

前記第 1 周波数信号は、前記第 1 音声入力部に入力された 1 フレーム分の前記第 1 音声信号を時間周波数変換することで取得され、前記第 2 周波数信号は、前記第 2 音声入力部に入力され、前記 1 フレーム分の第 1 音声信号に対応する、1 フレーム分の第 2 音声信号を時間周波数変換することで取得される、付記 1 のプログラム。

30

(付記 3)

前記所定周波数毎の位相差に基づいて、前記ユーザが前記第 1 音声入力部及び前記第 2 音声入力部の正面に位置する場合の発声方向位相差と、前記ユーザが前記第 1 音声入力部及び前記第 2 音声入力部の正面からずれた位置に存在する場合の発声方向位相差と、の差分に基づいて、前記第 1 閾値の値を定める、付記 1 または付記 2 のプログラム。

40

(付記 4)

第 2 閾値を超える周波数に対応する前記第 1 周波数信号及び第 2 周波数信号の位相差を使用して前記第 1 位相差外れ度合を算出する、付記 1 ~ 付記 3 の何れかのプログラム。

(付記 5)

前記第 1 周波数信号及び前記第 2 周波数信号の少なくとも一方の振幅が第 3 閾値を超える場合に、前記第 1 周波数信号及び第 2 周波数信号の位相差を使用して前記第 1 位相差外れ度合を算出する、

付記 1 ~ 付記 4 の何れかのプログラム。

50

(付記 6)

コンピュータが、

1 フレーム毎に、第 1 音声入力部に入力された第 1 音声信号に対応する第 1 周波数信号と第 2 音声入力部に入力された第 2 音声信号に対応する第 2 周波数信号との所定周波数毎の位相差に基づいて発声方向領域を決定し、前記所定周波数毎の位相差が前記発声方向領域から外れている第 1 位相差外れ度合を算出する処理を複数フレーム分実行し、前記複数フレーム分の前記第 1 位相差外れ度合から前記複数フレームの統計値である第 2 位相差外れ度合を算出し、前記第 2 位相差外れ度合が第 1 閾値以下である場合、ユーザが前記第 1 音声入力部及び前記第 2 音声入力部に向かって発声していると判定する、
発声方向判定方法。

10

(付記 7)

前記第 1 周波数信号は、前記第 1 音声入力部に入力された 1 フレーム分の前記第 1 音声信号を時間周波数変換することで取得され、
前記第 2 周波数信号は、前記第 2 音声入力部に入力され、前記 1 フレーム分の第 1 音声信号に対応する、1 フレーム分の第 2 音声信号を時間周波数変換することで取得される、
付記 6 の発声方向判定方法。

(付記 8)

前記所定周波数毎の位相差に基づいて、前記ユーザが前記第 1 音声入力部及び前記第 2 音声入力部の正面に位置する場合の発声方向位相差と、前記ユーザが前記第 1 音声入力部及び前記第 2 音声入力部の正面からずれた位置に存在する場合の発声方向位相差と、の差分に基づいて、前記第 1 閾値の値を定める、
付記 6 または付記 7 の発声方向判定方法。

20

(付記 9)

第 2 閾値を超える周波数に対応する前記第 1 周波数信号及び第 2 周波数信号の位相差を使用して前記第 1 位相差外れ度合を算出する、
付記 6 ~ 付記 8 の何れかの発声方向判定方法。

(付記 10)

前記第 1 周波数信号及び前記第 2 周波数信号の少なくとも一方の振幅が第 3 閾値を超える場合に、前記第 1 周波数信号及び第 2 周波数信号の位相差を使用して前記第 1 位相差外れ度合を算出する、
付記 6 ~ 付記 9 の何れかの発声方向判定方法。

30

(付記 11)

1 フレーム毎に、第 1 音声入力部に入力された第 1 音声信号に対応する第 1 周波数信号と第 2 音声入力部に入力された第 2 音声信号に対応する第 2 周波数信号との所定周波数毎の位相差に基づいて発声方向領域を決定し、前記所定周波数毎の位相差が前記発声方向領域から外れている第 1 位相差外れ度合を算出する処理を複数フレーム分実行する第 1 位相差外れ算出部と、
前記複数フレーム分の前記第 1 位相差外れ度合から前記複数フレームの統計値である第 2 位相差外れ度合を算出する第 2 位相差外れ算出部と、
前記第 2 位相差外れ度合が第 1 閾値以下である場合、ユーザが前記第 1 音声入力部及び前記第 2 音声入力部に向かって発声していると判定する発声方向判定部と、
を含む、発声方向判定装置。

40

(付記 12)

前記第 1 音声入力部に入力された 1 フレーム分の前記第 1 音声信号を時間周波数変換することで前記第 1 周波数信号を取得し、前記第 2 音声入力部に入力され、前記 1 フレーム分の第 1 音声信号に対応する、1 フレーム分の第 2 音声信号を時間周波数変換することで前記第 2 周波数信号を取得する、時間周波数変換部、
をさらに含む、付記 11 の発声方向判定装置。

(付記 13)

50

前記所定周波数毎の位相差に基づいて、前記ユーザが前記第 1 音声入力部及び前記第 2 音声入力部の正面に位置する場合の発声方向位相差と、前記ユーザが前記第 1 音声入力部及び前記第 2 音声入力部の正面からずれた位置に存在する場合の発声方向位相差と、の差分に基づいて、前記第 1 閾値の値を定める、

付記 1 1 または付記 1 2 の発声方向判定装置。

(付記 1 4)

前記第 1 位相差外れ算出部は、第 2 閾値を超える周波数に対応する前記第 1 周波数信号及び第 2 周波数信号の位相差を使用して前記第 1 位相差外れ度合を算出する、

付記 1 1 ~ 付記 1 3 の何れかの発声方向判定装置。

(付記 1 5)

前記第 1 位相差外れ算出部は、前記第 1 周波数信号及び前記第 2 周波数信号の少なくとも一方の振幅が第 3 閾値を超える場合に、前記第 1 周波数信号及び第 2 周波数信号の位相差を使用して前記第 1 位相差外れ度合を算出する、

付記 1 1 ~ 付記 1 4 の何れかの発声方向判定装置。

【符号の説明】

【 0 0 9 3 】

1 0 発声方向判定装置

2 1 A、2 1 B 音声入力部

2 2 時間周波数変換部

2 4 発声方向領域決定部

2 5 位相差外れ算出部

2 6 位相差外れ時間平均値算出部

2 7 発声方向判定部

5 1 C P U

5 2 一次記憶部

5 3 二次記憶部

5 5 A、5 5 B マイク

10

20

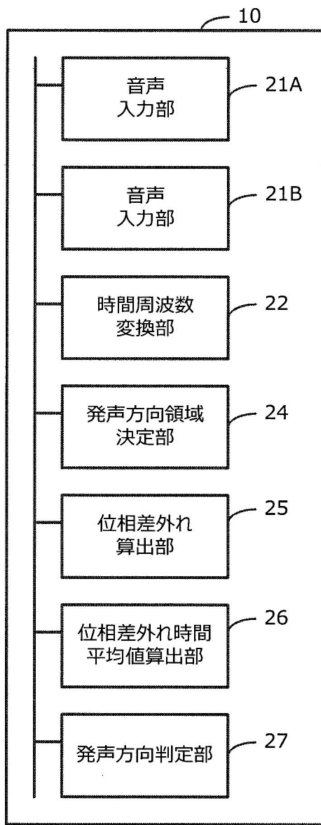
30

40

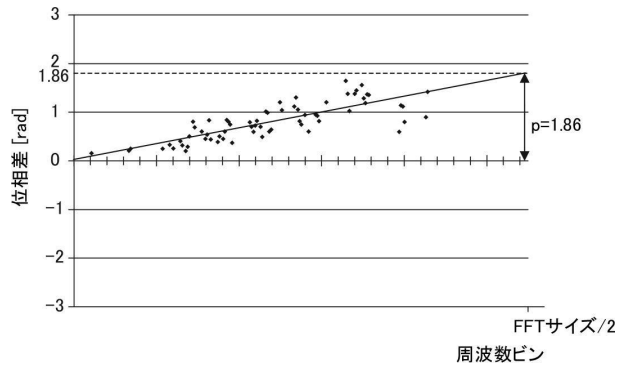
50

【図面】

【図 1】



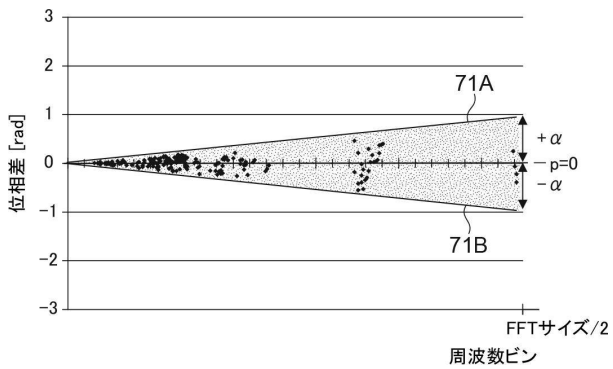
【図 2】



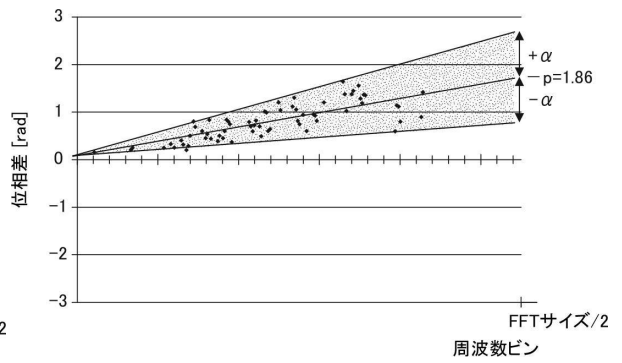
10

20

【図 3 A】



【図 3 B】

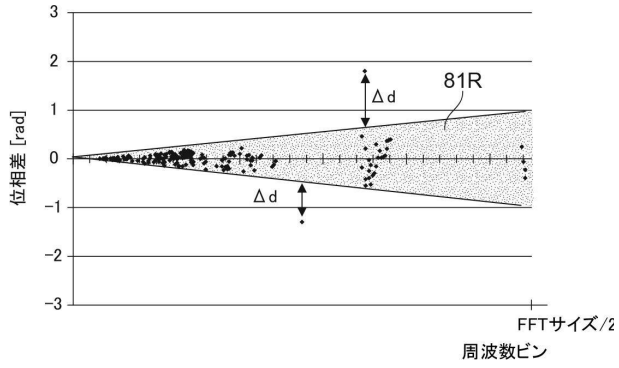


30

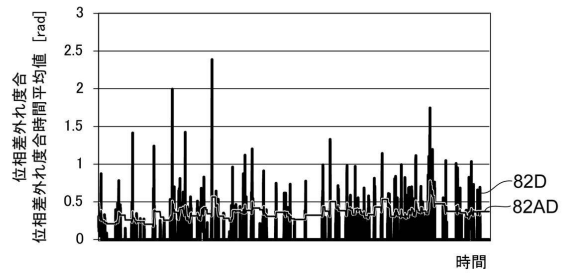
40

50

【図 4 A】

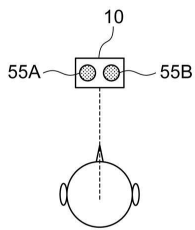


【図 4 B】

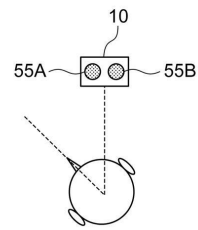


10

【図 5 A】

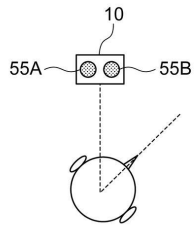


【図 5 B】

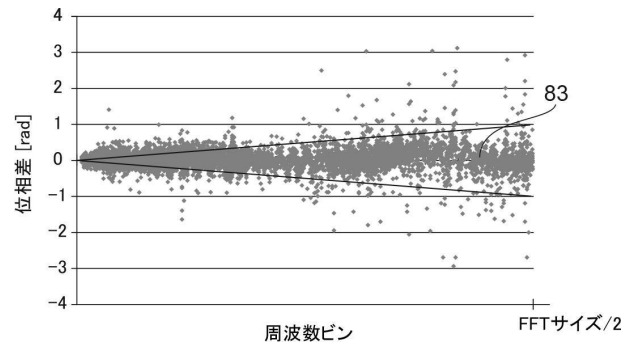


20

【図 5 C】



【図 6 A】

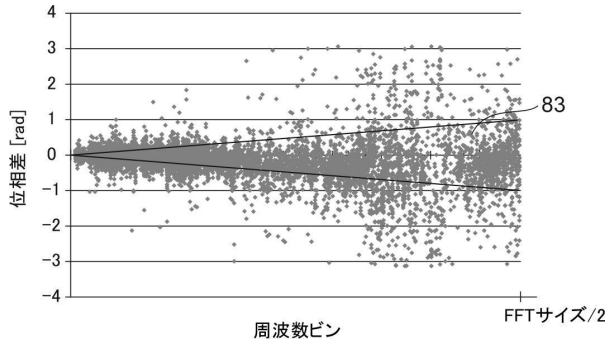


30

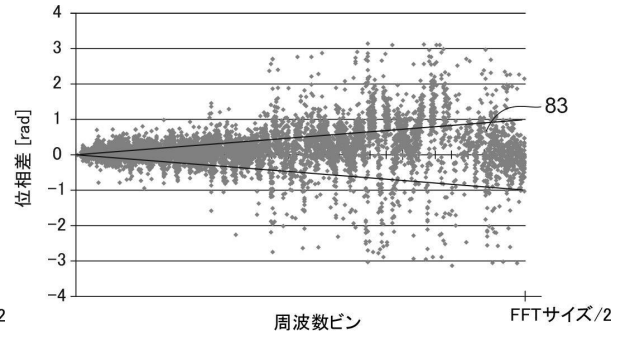
40

50

【図 6 B】

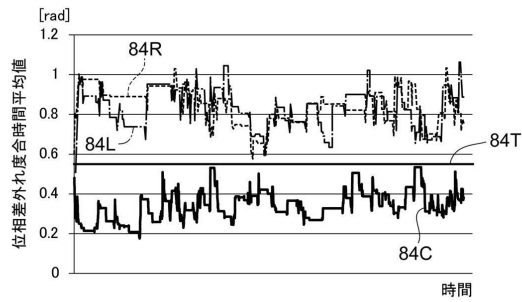


【図 6 C】

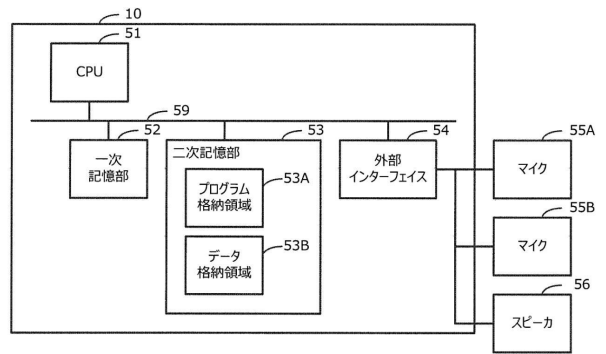


10

【図 7】



【図 8 A】



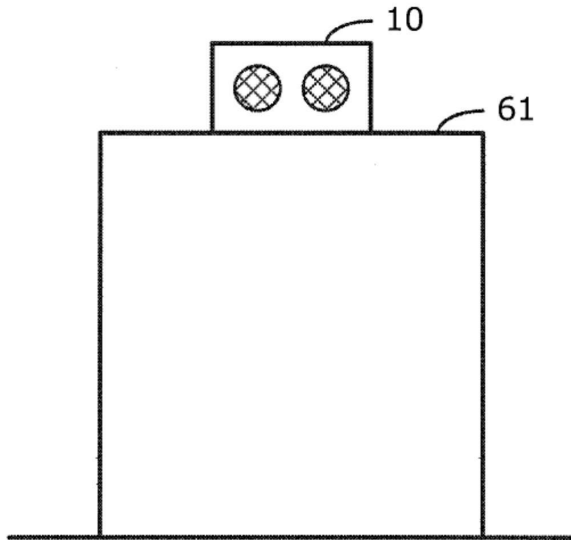
20

30

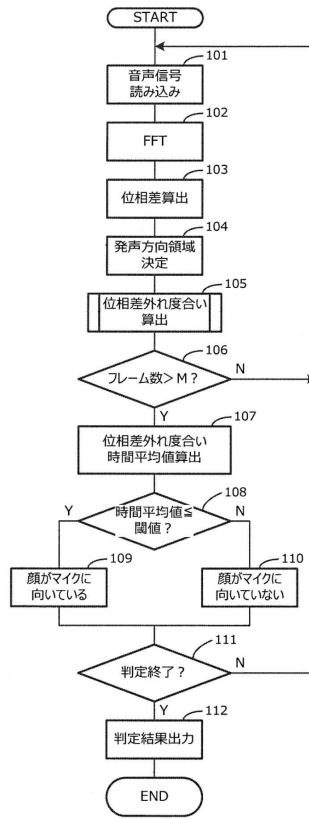
40

50

【図 8 B】



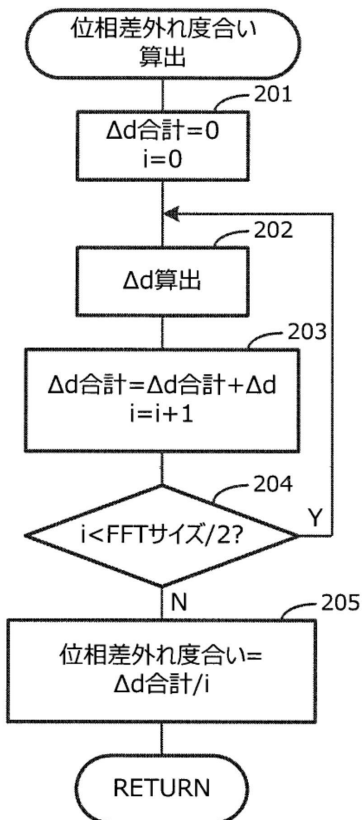
【図 9】



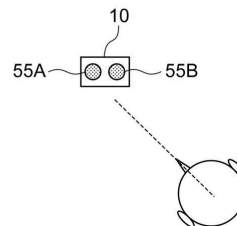
10

20

【図 10】



【図 11 A】

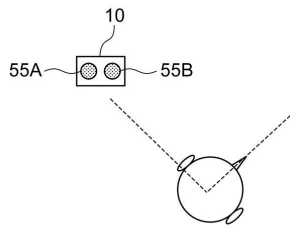


30

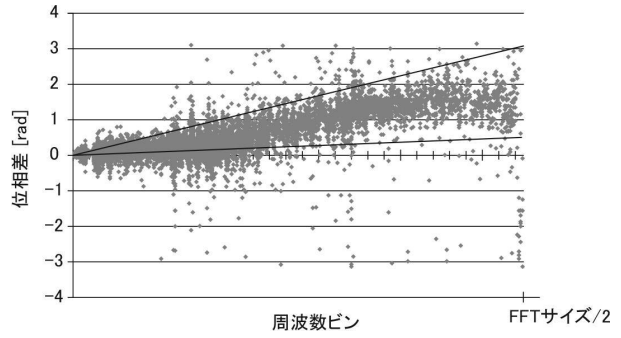
40

50

【図 1 1 B】

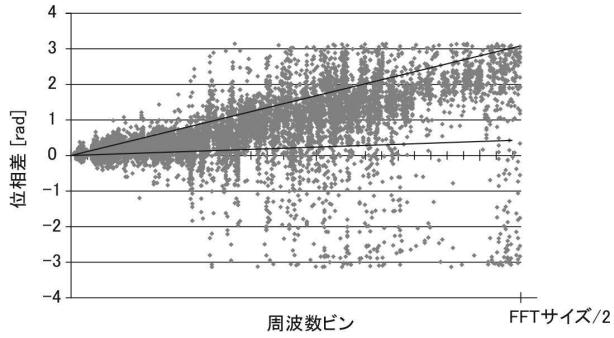


【図 1 2 A】

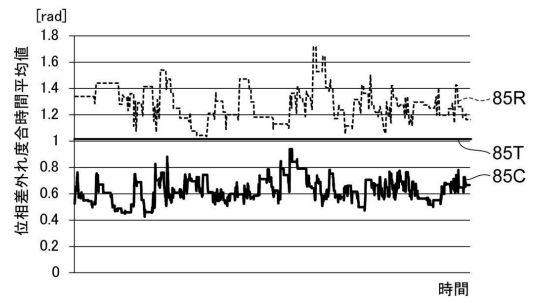


10

【図 1 2 B】

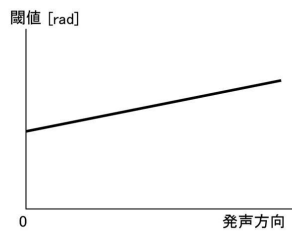


【図 1 3】

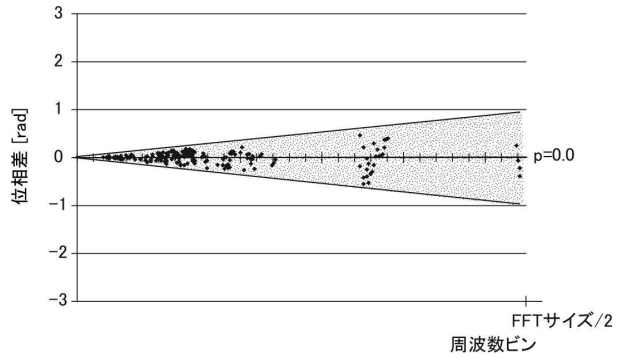


20

【図 1 4】



【図 1 5 A】

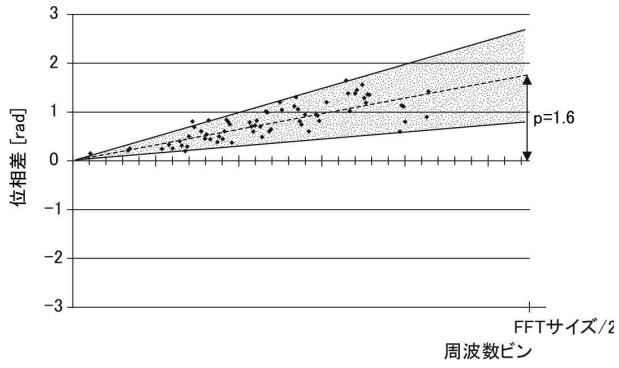


30

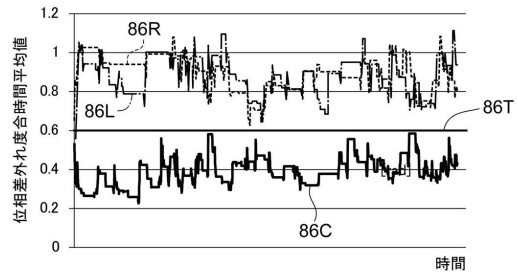
40

50

【図 15 B】

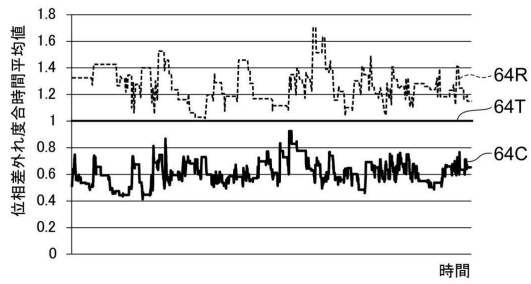


【図 16 A】

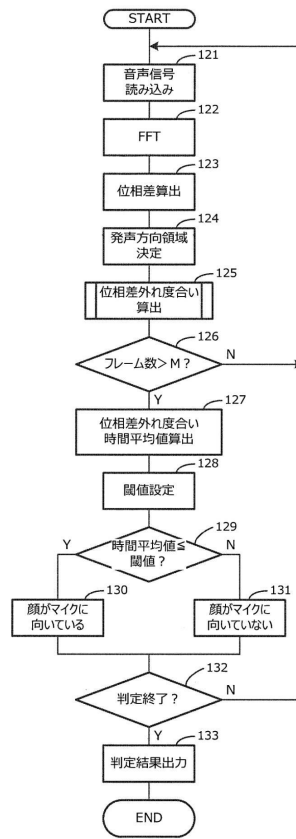


10

【図 16 B】



【図 17】



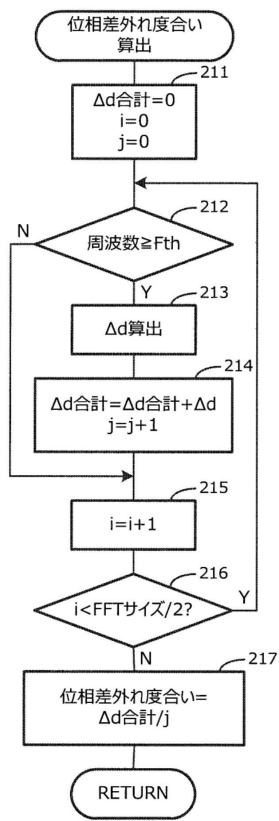
20

30

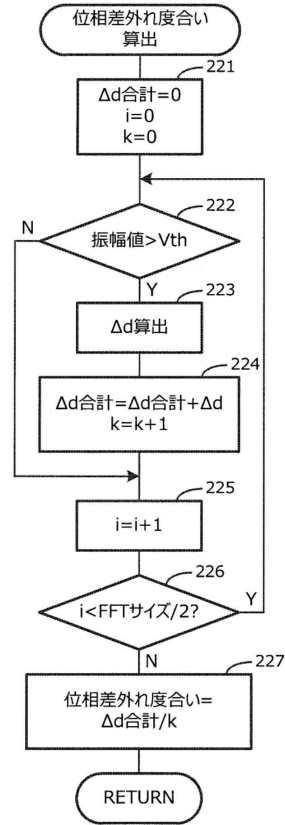
40

50

【図 18】



【図 19】



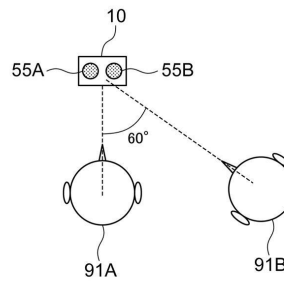
10

20

【図 20】

ユーザの位置 ユーザの顔の向き	ユーザの位置	
	マイクの正面	マイクの右側
マイクに顔を向ける	95%	93%
マイクに顔を向けない	92%	91%

【図 21 A】

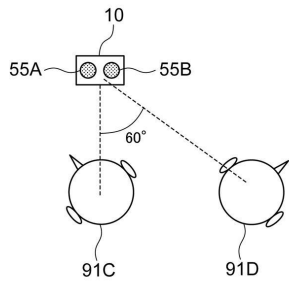


30

40

50

【 2 1 B 】



10

20

30

40

50

フロントページの続き

- (56)参考文献 特開2003-61184(JP,A)
特開2014-77736(JP,A)
特開2016-181789(JP,A)
特開2010-124447(JP,A)
特開2018-45195(JP,A)
米国特許出願公開第2012/0179458(US,A1)
- (58)調査した分野 (Int.Cl., DB名)
G10L 15/28, 25/00 - 25/93
H04R 1/40, 3/00