

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5493611号
(P5493611)

(45) 発行日 平成26年5月14日(2014.5.14)

(24) 登録日 平成26年3月14日(2014.3.14)

(51) Int.Cl.

H04R 3/00 (2006.01)
H04R 1/40 (2006.01)

F 1

H04R 3/00 320
H04R 1/40 320A

請求項の数 13 (全 16 頁)

(21) 出願番号 特願2009-207985 (P2009-207985)
 (22) 出願日 平成21年9月9日 (2009.9.9)
 (65) 公開番号 特開2011-61422 (P2011-61422A)
 (43) 公開日 平成23年3月24日 (2011.3.24)
 審査請求日 平成24年8月28日 (2012.8.28)

(73) 特許権者 000002185
 ソニー株式会社
 東京都港区港南1丁目7番1号
 (74) 代理人 100095957
 弁理士 龟谷 美明
 (74) 代理人 100096389
 弁理士 金本 哲男
 (74) 代理人 100101557
 弁理士 萩原 康司
 (74) 代理人 100128587
 弁理士 松本 一騎
 (72) 発明者 千原 秀一
 東京都港区港南1丁目7番1号 ソニー株式会社内

最終頁に続く

(54) 【発明の名称】情報処理装置、情報処理方法およびプログラム

(57) 【特許請求の範囲】

【請求項 1】

ユーザーの頭部の左右に位置するユニットに少なくとも対をなして設けられ、当該ユーザーが発話した音声を少なくとも外部の音声を収音して音声信号に変換する収音部と、

少なくともユーザーの指示に応じて、少なくとも前記収音部の感度を規定する処理パラメータを設定するパラメータ設定部と、

前記処理パラメータに基づいて、前記収音部から入力される音声信号にビームフォーミング処理を含む音声処理を施す音声信号処理部と、

を備え、

前記処理パラメータは、当該ユーザーの指示に応じて隨時設定可能であり、

前記ビームフォーミング処理においては、音源と一対の前記収音部の各々との距離の差に応じた、当該収音部の各々によって収音される音声の位相差を、所定のしきい値と比較することにより、前記ユーザーが発話した音声と他の音声とが判別される、情報処理装置。

【請求項 2】

前記音声信号処理部は、前記処理パラメータに基づいて、前記収音部間の感度バランスを調整する、請求項 1 に記載の情報処理装置。

【請求項 3】

前記音声信号処理部は、前記処理パラメータに基づいて、前記収音部の感度を調整する

、請求項 1 に記載の情報処理装置。

【請求項 4】

前記音声信号処理部は、前記処理パラメータに基づいて、前記収音部から入力される音声信号のレベルが継続して所定の閾値未満である場合に、音声信号の入力を中止するまでの継続時間を調整する、請求項 1 に記載の情報処理装置。

【請求項 5】

前記音声信号処理部は、前記処理パラメータに基づいて、前記収音部から入力される音声信号の周波数範囲を調整する、請求項 1 に記載の情報処理装置。

【請求項 6】

前記収音部と特定音源の相対的な位置変化に追従して、前記特定音源に対応する音声信号のレベルが最大となるように、前記収音部間の感度バランスを自動設定する、請求項 1 に記載の情報処理装置。 10

【請求項 7】

前記音声処理を施された音声信号を通信網を介して再生装置に送信する送信部と、

前記処理パラメータを指定するパラメータ指定情報を前記再生装置から受信する受信部と、

をさらに備え、

前記パラメータ設定部は、前記受信されたパラメータ指定情報に従って、前記処理パラメータを設定する、請求項 1 に記載の情報処理装置。

【請求項 8】

前記音声信号処理部は、前記収音部の各々から入力される音声信号の位相差が所定の閾値未満である場合に、音声信号のレベルを維持し、所定の閾値以上である場合に、音声信号のレベルを減少させる、請求項 1 に記載の情報処理装置。 20

【請求項 9】

前記音声信号処理部は、前記収音部から入力される音声信号のうち特定音源以外に対応する音声信号以外の信号を除去するための信号を、前記収音部から入力される音声信号に合成する、請求項 1 に記載の情報処理装置。

【請求項 10】

前記収音部は、ヘッドホンの左右ユニットに対をなして設けられる、請求項 1 に記載の情報処理装置。 30

【請求項 11】

前記音声信号処理部は、前記処理パラメータを設定するための設定画面を通じて入力されるユーザーの指示に応じて、前記処理パラメータを調整する、請求項 1 に記載の情報処理装置。

【請求項 12】

ユーザーの頭部の左右に位置するユニットに少なくとも対をなして設けられ、当該ユーザーが発話した音声を少なくとも含む外部の音声を収音して音声信号に変換する収音部の感度を規定する処理パラメータを、少なくともユーザーの指示に応じて設定することと、

前記処理パラメータに基づいて、前記音声信号にビームフォーミング処理を含む音声処理を施すことと、 40

を含み、

前記処理パラメータは、当該ユーザーの指示に応じて隨時設定可能であり、

前記ビームフォーミング処理においては、音源と一対の前記収音部の各々との距離の差に応じた、当該収音部の各々によって収音される音声の位相差を、所定のしきい値と比較することにより、前記ユーザーが発話した音声と他の音声とが判別される、情報処理方法。

【請求項 13】

ユーザーの頭部の左右に位置するユニットに少なくとも対をなして設けられ、当該ユーザーが発話した音声を少なくとも含む外部の音声を収音して音声信号に変換する収音部の感度を規定する処理パラメータを、少なくともユーザーの指示に応じて設定することと、 50

前記処理パラメータに基づいて、前記音声信号にビームフォーミング処理を含む音声処理を施すことと、
を含み、

前記処理パラメータは、当該ユーザーの指示に応じて隨時設定可能であり、
前記ビームフォーミング処理においては、音源と一対の前記収音部の各々との距離の差
に応じた、当該収音部の各々によって収音される音声の位相差を、所定のしきい値と比較
することにより、前記ユーザーが発話した音声と他の音声とが判別される、情報処理方法
をコンピュータに実行させるプログラム。

10

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、情報処理装置、情報処理方法およびプログラムに関する。

【背景技術】

【0002】

Voice over Internet Protocol (VoIP) を用いた IP 電話システム、会議システム等の音声処理システムでは、遠隔地に送信する送信音声の入力にビームフォーミングが用いられる場合がある。この場合、ビームフォーミングに対応するマイクアレイを用いて、特定方向からの音声が送信音声として選択的に入力される。これにより、発話者および発話者と同一線上にある音源の音声（以下、特定音声とも称する。）を維持する一方で、環境音（ノイズ）である不特定音源の音声（以下、不特定音声とも称する。）を弱めることで、送信音声を良好な状態で入力することができる。

20

【先行技術文献】

【特許文献】

【0003】

【特許文献1】特開平06-233388号公報

【発明の概要】

【発明が解決しようとする課題】

【0004】

30

ビームフォーミングでは、マイクアレイの各マイクにより収音された音声が音声間の位相差、音量差等に基づいて処理される。このため、送信音声の品質は、各マイク間の感度バランスの差、各マイクの感度自体のバラツキ、入力音声の周波数範囲等、各種の処理パラメータの影響を受ける。

【0005】

しかし、従来、処理パラメータの変更には回路的な調整等が要求されるため、ユーザーは、使用環境に応じて処理パラメータを設定して、送信音声の品質を向上させることが困難であった。

【0006】

そこで、本発明は、ビームフォーミングを用いて入力される送信音声の品質を向上可能な、情報処理装置、情報処理方法およびプログラムを提供しようとするものである。

40

【課題を解決するための手段】

【0007】

本発明のある実施形態によれば、少なくとも一対で設けられ、外部の音声を収音して音声信号に変換する収音部と、少なくともユーザーの指示に応じて、少なくとも収音部の感度を規定する処理パラメータを設定するパラメータ設定部と、処理パラメータに基づいて、収音部から入力される音声信号にビームフォーミング処理を含む処理を施す音声信号処理部と、を備える情報処理装置が提供される。

【0008】

かかる構成によれば、少なくとも一対で設けられる収音部により収音された外部の音声

50

信号に、少なくとも収音部の感度を規定し、少なくともユーザーの指示に応じて設定された処理パラメータに基づいて、ビームフォーミング処理を含む音声処理が施される。これにより、使用環境に応じて、少なくとも収音部の感度を規定する処理パラメータを設定することで、特定音声が良好な状態で選択的に入力可能となり、送信音声の品質を向上させることができる。

【0009】

また、本発明の別の実施形態によれば、少なくともユーザーの指示に応じて、音声信号の処理条件を規定する処理パラメータを設定するステップと、少なくとも一対で設けられる収音部から入力される外部の音声信号に、処理パラメータに基づいて、ビームフォーミング処理を含む音声処理を施すステップと、を含む情報処理方法が提供される。

10

【0010】

また、本発明の別の実施形態によれば、上記情報処理方法をコンピュータに実行させるためのプログラムが提供される。プログラムは、コンピュータ読取り可能な記録媒体を用いて提供されてもよく、通信手段を介して提供されてもよい。

【発明の効果】

【0011】

本発明によれば、ビームフォーミングを用いて入力される送信音声の品質を向上可能な、情報処理装置、情報処理方法およびプログラムを提供することができる。

【図面の簡単な説明】

【0012】

20

【図1】ビームフォーミングの原理を示す図である。

【図2】ビームフォーミングに用いられる音声間の位相差の算定方法を示す図である。

【図3】情報処理装置の主要なハードウェア構成例を示す図である。

【図4】音声信号処理部の主要な機能構成を示す図である。

【図5】処理パラメータ設定用の設定パネルを示す図である。

【図6A】感度バランス調整の設定処理を説明する図(1/2)である。

【図6B】感度バランス調整の設定処理を説明する図(2/2)である。

【図7A】感度調整の設定処理を説明する図(1/2)である。

【図7B】感度調整の設定処理を説明する図(2/2)である。

【図8A】感度調整補正の設定処理を説明する図(1/2)である。

30

【図8B】感度調整補正の設定処理を説明する図(2/2)である。

【図9】周波数調整の設定処理を説明する図である。

【図10A】特定音源の追跡処理を説明する図(1/2)である。

【図10B】特定音源の追跡処理を説明する図(2/2)である。

【図11】処理パラメータの遠隔設定処理を説明する図である。

【発明を実施するための形態】

【0013】

以下に添付図面を参照しながら、本発明の好適な実施の形態について詳細に説明する。なお、本明細書および図面において、実質的に同一の機能構成を有する構成要素については、同一の符号を付することにより重複説明を省略する。

40

【0014】

[1. ビームフォーミング]

まず、図1および図2を参照しながら、ビームフォーミングの原理について説明する。図1は、ビームフォーミングの原理を示す図である。図2は、ビームフォーミングに用いられる音声間の位相差の算定方法を示す図である。

【0015】

図1には、発話者Uが装着するヘッドホンHPの左右ユニットに、マイクアレイを構成する一対の無指向性マイクM1、M2を設ける場合が示されている。なお、マイクM1、M2は、ヘッドホンHPに限定されず、ヘッドバンドの左右ユニット、帽子の左右等に設けられてもよく、2以上で設けられてもよい。

50

【0016】

発話者UがヘッドホンHPを装着した状態で発話すると、マイクM1、M2から略等距離に位置する発話者Uの口元を特定音源Ssとして、発話者Uの音声(特定音声Vs)がマイクM1、M2により略同時に、略同音量かつ略同位相で収音される。一方、ノイズ等の環境音(不特定音声Vn)は、概してマイクM1、M2から異なる距離に位置する不特定音源Snから発せられるので、マイクM1、M2により異なる時点、異なる音量かつ異なる位相で収音される。特に、ヘッドホンHPにマイクM1、M2を設ける場合、発話者Uが動作等しても、マイクM1、M2から略等距離の位置に特定音源Ssが位置しているので、特定音声Vsと不特定音声Vnを容易に判別することができる。

【0017】

ここで、マイクM1、M2により収音される音声V間の位相差は、図2を用いて算定される。音源SとマイクM1、M2の距離SM1、SM2が次式により得られる。

$$SM1 = ((L \cdot \tan + d)^2 + L^2)$$

$$SM2 = ((L \cdot \tan - d)^2 + L^2)$$

d : マイクM1、M2間の距離の1/2

L : 音源Sとマイクアレイ間の垂直距離

: 音源Sとマイクアレイ中心との角度

よって、マイクM1、M2の音声V間の位相差が次式により得られる。

$$= 2 f \cdot (SM1 - SM2) / c$$

c : 音速(342m/s)

f : 音声の周波数(Hz)

【0018】

ビームフォーミングでは、マイクM1、M2により収音される音声Vの位相差等に基づいて、特定音声Vsを維持する一方で、不特定音声Vnを弱めることで、特定音声Vsを送信音声として選択的に入力することができる。

【0019】

マイクM1、M2により収音される音声Vは、音声Vの位相差を閾値tと比較することで、特定音声Vsまたは不特定音声Vnとして判別される。例えばd = 5cm、L = 100cm、f = 800Hzの場合、位相差 = 42°を閾値tとして、閾値t未満の音声Vが特定音声Vsとして、閾値t以上の音声Vが不特定音声Vnとして判別される。ここで、判別に用いる閾値tは、d、L等の条件に応じて異なる値となる。なお、閾値tは、絶対値が同一の正值/負値として定義されるが、以下では、| | < tを閾値t未満、| | | を閾値t以上と各々に称する。

【0020】

[2. 情報処理装置100の構成]

次に、図3および図4を参照しながら、本発明の一実施形態に係る情報処理装置100について説明する。図3は、情報処理装置100の主要なハードウェア構成例を示す図である。図4は、音声信号処理部150の主要な機能構成を示す図である。

【0021】

図3に示すように、情報処理装置100は、例えば、パーソナルコンピュータ、PDA、ゲーム装置、携帯電話等であるが、以下では、情報処理装置100がパーソナルコンピュータである場合を想定する。

【0022】

情報処理装置100は、主に、CPU101、ROM103、RAM105、ホストバス107、ブリッジ109、外部バス111、インターフェース113、音声入出力装置115、操作装置117、表示装置119、ストレージ装置121、ドライブ123、接続ポート125、通信装置127を含んで構成される。

【0023】

CPU101は、演算処理装置および制御装置として機能し、ROM103、RAM105、ストレージ装置121、またはリムーバブル記録媒体129に記録された各種プロ

10

20

30

40

50

グラムに従って、情報処理装置 100 の動作を少なくとも部分的に制御する。CPU101 は、少なくともユーザーの指示に応じて、音声信号の処理条件を規定する処理パラメータを設定するパラメータ設定部としても機能する。ROM103 は、CPU101 が用いるプログラムやパラメータ等を記憶する。RAM105 は、CPU101 が実行するプログラム、プログラム実行時のパラメータ等を一時記憶する。

【0024】

CPU101、ROM103、RAM105 は、ホストバス 107 により互いに接続される。ホストバス 107 は、ブリッジ 109 を介して外部バス 111 に接続される。

【0025】

音声入出力装置 115 は、ヘッドホン HP、マイク、スピーカ等を含む、音声信号を入出力可能な入出力手段である。音声入出力装置 115 は、各種フィルタ 181、185、A/D 変換器 183、D/A 変換器（不図示）等の前処理部 116 を含む（図 4 参照）。特に、本実施形態に係る音声入出力装置 115 では、ヘッドホン HP の左右ユニットに一对のマイク M1、M2 が設けられている。音声入出力装置 115 は、マイク M1、M2 により収音された外部の音声信号を音声信号処理部 150 に供給し、音声信号処理部 150 により処理された音声信号をヘッドホン HP に供給する。

【0026】

操作装置 117 は、例えば、マウス、キーボード、タッチパネル、ボタン、スイッチ等、ユーザーが操作可能な操作手段である。操作装置 117 は、例えば上記の操作手段を用いてユーザーにより入力された操作情報に基づいて入力信号を生成し、CPU101 に出力する入力制御回路等を含んで構成される。ユーザーは、操作装置 117 の操作を介して、情報処理装置 100 に対して各種のデータを入力し、処理動作を指示する。

【0027】

表示装置 119 は、例えば、液晶ディスプレイ等の表示手段である。表示装置 119 は、情報処理装置 100 の処理結果を出力する。例えば、表示装置 119 は、後述する各種パラメータ設定用の設定パネル CP を含む情報処理装置 100 による処理結果を、テキスト情報または画像情報として表示する。

【0028】

ストレージ装置 121 は、データ格納用の装置であり、例えば、HDD 等の磁気記憶デバイス等を含む。ストレージ装置 121 は、CPU101 が実行するプログラム、各種データ、外部から取得された各種データ等を格納する。

【0029】

ドライブ 123 は、記録媒体用リーダライタであり、情報処理装置 100 に内蔵または外付けされる。ドライブ 123 は、装着される磁気ディスク等のリムーバブル記録媒体 129 に対して、記録済みデータを読み出して RAM105 に出力し、記録対象のデータを書き込む。

【0030】

接続ポート 125 は、例えば、USB ポート等、外部機器 131 を情報処理装置 100 に直接接続するためのポートである。情報処理装置 100 は、接続ポート 125 に接続された外部機器 131 に対して、接続ポート 125 を介してデータを取得し、データを提供する。

【0031】

通信装置 127 は、通信網 N に接続するための通信デバイス等から構成される通信インターフェース 113 である。通信装置 127 は、例えば、有線または無線 LAN 用の通信カード等である。通信装置 127 に接続される通信網 N は、有線または無線により接続されたネットワーク等により構成される。

【0032】

[3. 音声信号処理部 150 の構成]

図 4 に示すように、情報処理装置 100 は、マイク M1、M2 の音声信号を処理する音声信号処理部 150 を含む。音声信号処理部 150 は、ハードウェア、ソフトウェア、ま

10

20

30

40

50

たは両者の組合せにより実現される。なお、図4には、本発明に関する音声入力処理を行うための構成のみが示されている。

【0033】

音声信号処理部150は、マイクM1、M2の入力系統毎に、感度調整部151、感度調整補正部153、および周波数調整部155を含む。また、音声信号処理部150は、マイクM1、M2の入力系統の後段に、時間差分析部157、周波数分析部159、位相差分析部161、ビームフォーミング処理部163（B.F.処理部163とも称する。）、ノイズ生成部165、ノイズ除去部167、および加算器169を含む。なお、ノイズ除去処理を行わない場合、ノイズ生成部165、ノイズ除去部167、および加算器169が省略されてもよい。

10

【0034】

マイクM1、M2は、外部の音声を収音してアナログの音声信号に変換し、前処理部116に供給する。前処理部116では、マイクM1、M2の音声信号がフィルタ181に入力される。フィルタ181は、音声信号に含まれる所定の信号成分をフィルタリングし、A/D変換器183に供給する。A/D変換器183は、フィルタリング後の音声信号をデジタルの音声信号（音声データ）にPCM変換し、音声信号処理部150に供給する。

【0035】

音声信号処理部150では、マイクM1、M2の入力系統毎に、感度調整部151、感度調整補正部153、および周波数調整部155による信号処理が施され、時間差分析部157および周波数分析部159に供給される。なお、感度調整部151、感度調整補正部153、および周波数調整部155による信号処理の詳細については、後述する。

20

【0036】

時間差分析部157は、各入力系統から供給される音声信号に基づいて、各マイクM1、M2に到達する音声の時間差を分析する。音声の到達時間差は、各マイクM1、M2の音声信号の時系列を対象として、例えば、位相変化、レベル変化等に基づく相互相關分析を行うことで分析される。

【0037】

周波数分析部159は、各入力系統から供給される音声信号に基づいて、音声信号の周波数を分析する。周波数分析では、FFT（高速フーリエ変換）等を用いて音声信号の時系列を様々な周期・振幅の正弦波信号に分解し、音声信号の周波数スペクトルを分析する。

30

【0038】

位相差分析部161は、時間差分析および周波数分析の結果に基づいて、各マイクM1、M2により収音された音声間の位相差を分析する。位相差分析では、周波数成分毎に音声の位相差が分析される。位相差分析により、周波数成分毎の位相差を所定の閾値tと比較し、閾値t以上の周波数成分をノイズ成分（不特定音声vn）として判別することができる。

【0039】

B.F.処理部163は、位相差分析の結果に基づいて、各入力系統から供給される音声信号にビームフォーミング処理を施して加算器169に供給する。ビームフォーミング処理では、各マイクM1、M2により収音された音声間の位相差が閾値t未満である場合に、信号レベルが維持され、閾値t以上である場合に、信号レベルが減少される。

40

【0040】

これにより、特定音声Vsは、マイクM1、M2から略等距離の位置を音源Ssとしており、位相差が小さいので、信号レベルが維持される。一方、不特定音声vnは、概してマイクM1、M2から異なる距離の位置を音源Snとしており、位相差が大きいので、信号レベルが減少される。

【0041】

ノイズ生成部165は、位相差分析の結果に基づいて、マイクM1、M2により収音さ

50

れた音声に含まれるノイズ(不特定音声Vn)を表すノイズ信号を生成する。

【0042】

ノイズ除去部167は、不特定音声Vnに相当する信号成分を除去するために、ノイズ信号の反転により表される信号を生成して加算器169に供給する。ここで、ノイズ除去部167は、加算処理後の音声信号をフィードバックされ、フィードバック信号にノイズ信号を適応させる。

【0043】

加算器169は、BF処理部163から供給される音声信号とノイズ除去部167から供給される信号を合算してフィルタ185に供給する。これにより、BF処理後の音声信号からノイズ成分が除去され、特定音声がさらに選択的に入力されるようになる。合算後の音声信号は、後段のフィルタ185を介して送信音声として入力され、通信装置127により通信網Nを介して不図示の再生装置100'に送信されて再生される。10

【0044】

[4. 処理パラメータの設定処理]

次に、図5～図11を参照しながら、処理パラメータの設定処理について説明する。図5は、処理パラメータ設定用の設定パネルCPを示す図である。図6A、6Bおよび図7A、7Bは、感度バランス調整および感度調整の設定処理を各々に説明する図である。図8A、8Bおよび図9は、感度調整補正および周波数調整の設定処理を各々に説明する図である。図10A、10Bおよび図11は、特定音源Ssの追跡処理、および処理パラメータの遠隔設定処理を各々に説明する図である。20

【0045】

処理パラメータの設定に際して、CPU101は、プログラムの実行により図5に示すような設定パネルCPを表示装置119に表示させる。設定パネルCPには、感度バランス調整、感度調整、感度調整補正、周波数調整の各パラメータを設定するためのスライダC1、C2、C3、C4が表示されている。また、設定パネルCPには、音源追跡処理、遠隔設定処理の有効／無効を切替えるためのスイッチC5、C6とともに、レベルメータLMが表示されている。なお、設定パネルCPに表示される操作用アイコンは、スライダ、スイッチ以外のアイコンでもよい。

【0046】

感度バランス調整用のスライダC1では、ノブI1の操作によりパラメータが設定される。感度調整、感度調整補正、周波数調整用のスライダC2、C3、C4では、ノブI21、I22、I31、I32、I41、I42、I43、I44の操作により、マイクM1、M2毎にパラメータが設定される。なお、感度調整、感度調整補正、周波数調整用のスライダC2、C3、C4は、マイクM1、M2毎に設けられる代わりに、マイクM1、M2に共通して設けられてもよい。レベルメータLMには、マイクM1、M2毎に特定音声Vsおよび不特定音声Vnの信号レベルL1～L4が表示される。30

【0047】

発話者Uは、所定の操作により設定パネルCPを表示させ、設定パネルCP上でスライダC1～C4およびスイッチC5、C6を操作して、各パラメータおよびモードを設定することができる。40

【0048】

[4-1. 感度バランス調整処理]

感度調整部151は、感度バランス調整パラメータに基づいて、マイクM1、M2の信号間のレベルバランスを変化させて、マイクM1、M2間の感度バランスを調整する。

【0049】

装着用のマイクM1、M2の感度には、製造条件により、+/-3dB程度のバラツキが生じることが知られている。例えば、音量差のパラメータを用いて音源位置の特定精度を向上させるアルゴリズムを適用する場合等を想定する。この場合、マイクM1、M2に感度差が存在すると、収音される音声の音量に差が生じ、発話者Uの正面に位置する音源の音声が発話者Uの正面からずれて位置する音源の音声として収音されてしまう。また、50

同一感度のマイク M 1、M 2 を用いることも考えられるが、マイク部品の製造歩留まりが低下し、コスト増加の要因となってしまう。

【 0 0 5 0 】

例えば、図 6 A に示すように、マイク M 1 の感度がマイク M 2 よりも高い場合、マイク M 1 の信号レベルが相対的に高くなる。よって、例えば、発話者 U の正面に位置する音源 S s の特定音声 V s は、マイク M 1 の側に位置する音源 S s' の音声 V s' として収音されてしまう。そして、特定音源 S s の音声は、受話者 U' により音源 S s' の音声 V s' として聴取されてしまう。

【 0 0 5 1 】

この場合、図 6 B に示すように、感度バランス調整用スライダ C 1 を用いて、マイク M 1、M 2 の信号間のレベルバランスがマイク M 2 の側にシフトするように、感度バランス調整パラメータが設定される。ここで、レベルバランスのシフトは、マイク M 2 の信号レベルの増加、マイク M 1 の信号レベルの減少、または（例えばマイク M 1、M 2 の信号レベルの合計が調整前後で変化しないような）両者の組合せにより実現される。例えばマイク M 2 の信号レベルを増加する場合、マイク M 2 の信号レベルに所定の増加率が乗算され、マイク M 1、M 2 間で信号レベル差が低減される。これにより、感度バランスのバラツキ等に拘らずに、特定音源 S s の音声を発話者 U の正面に位置する音源の音声として入力することができる。

【 0 0 5 2 】

[4 - 2 . 感度調整処理]

また、感度調整部 151 は、感度調整パラメータに基づいて、マイク M 1、M 2 の信号レベルを変化させて、マイク M 1、M 2 の感度を調整する。マイクの感度を上げると、マイクから離れた音源の音声が入力可能となるが、不特定音声 V n も入力され易くなる。一方、マイクの感度を下げるとき、マイクに近い音源の音声のみが入力可能となり、特定音声 V s を選択的に入力し易くなる。

【 0 0 5 3 】

また、感度調整では、特定音声 V s および不特定音声 V n について、信号レベルをリアルタイムに表示するレベルメータ L M が利用される。レベルメータ L M は、周波数分析された信号レベルをリアルタイムに表示することで実現される。一般に、送信音声が受話者 U' の側でしか再生されないので、発話者 U は、感度調整の結果を容易に確認することができない。しかし、レベルメータ L M を用いることで、特定音声 V s と不特定音声 V n の入力状況が確認可能となり、感度調整を容易に行うことができる。

【 0 0 5 4 】

図 7 A に示す例では、マイク M 1、M 2 の感度が高いので、特定音声 V s とともに、不特定音声 V n が相当程度で入力されている。ここで、発話者 U は、レベルメータ L M を通じて音声の入力状況（L 1、L 3 : V s の入力状況、L 2、L 4 : V n の入力状況）を確認することができる。

【 0 0 5 5 】

この場合、図 7 B に示すように、感度調整用スライダ C 2 を用いて、マイク M 1、M 2 の感度を低下させるように、感度調整パラメータが設定される（なお、図 7 A、7 B 中では、マイク M 1 のスライダのみが示されている。）。そして、マイク M 1、M 2 の信号レベルに、感度調整パラメータの設定に応じて所定の低減率が乗算され、マイク M 1、M 2 の信号レベルが低減される。ここで、発話者 U は、レベルメータ L M を通じて音声の入力状況を確認しながら、感度調整を適切に行うことで、特定音声 V s を良好な状態で選択的に入力することができる。

【 0 0 5 6 】

[4 - 3 . 感度調整補正処理]

感度調整補正部 153 は、感度調整補正パラメータに基づいて、マイク M 1、M 2 の感度調整を補正する。ここで、感度調整補正パラメータは、信号レベルが継続して所定の閾値 L t 未満である場合に、音声信号の入力を中止するまでの継続時間 t t を示すパラメー

10

20

30

40

50

タである。ここで、所定の閾値 L_t は、マイク M1、M2 の感度調整結果に応じて設定される。

【0057】

発話音声は、一定の音量で継続するものではない。よって、特定音声 V_s の音量が一時的に下がると、低い音量の音声が入力されず、特定音声 V_s が断続的に入力されてしまう。しかし、マイクの感度を上げ過ぎると、低い音量の不特定音声 V_n も入力され、信号ノイズ比 (S/N) が低下してしまう。

【0058】

このため、感度調整補正部 153 は、所定の閾値 L_t 未満の信号レベルが検出されると、音声信号の入力を中止するか否かの判定を開始する。そして、判定時間 t_t に亘って所定の閾値 L_t 未満の信号レベルが検出された場合に、音声信号の入力を中止する。一方、判定時間 t_t 内に所定の閾値 L_t 以上の信号レベルが再び検出された場合に、判定時間 t_t を初期化し、音声信号の入力を継続する。10

【0059】

図 8 A に示す例では、信号レベルが所定の閾値 L_t を境として上下に変動している。また、閾値 L_t 未満の区間長 t_t が継続時間 t_t 以上となっている。このため、閾値 L_t 未満の区間の音声信号が入力されず、特定音声 V_s が断続的に入力されている。

【0060】

この場合、図 8 B に示すように、感度調整補正用スライダ C_3 を用いて、継続時間 t_t が長くなるように、感度調整補正パラメータが設定される（なお、図 8 A、8 B 中では、マイク M1 のスライダのみが示されている。）。これにより、閾値 L_t 未満の区間の音声信号が入力され、特定音声 V_s を継続的に入力することができる。20

【0061】

[4 - 4 . 周波数調整処理]

周波数調整部 155 は、周波数調整パラメータに基づいて、各マイク M1、M2 から入力される音声信号の周波数範囲を調整する。固定電話では、発話音声の周波数帯域として 300 ~ 3400 Hz 程度が利用されている。一方、環境音（ノイズ）の周波数帯域は、発話音声の周波数帯域よりも広いことが知られている。

【0062】

このため、図 9 に示すように、周波数調整用スライダ C_4 を用いて、入力される音声信号の周波数範囲が設定される。ここで、周波数範囲は、周波数範囲の上限および下限を各々に示すタブ I_4_1、I_4_2 を操作することで設定される（なお、図 9 では、マイク M1 のスライダのみが示されている。）。周波数調整部 155 は、設定された周波数範囲に基づいて、音声信号から所定の信号成分をフィルタリングして後段に供給する。これにより、特定音声 V_s を良好な状態で選択的に入力することができる。30

【0063】

[4 - 5 . 音源追跡処理]

音源追跡処理では、マイク M1、M2 と特定音源 S_s の相対的な位置変化に追従して、感度バランス調整パラメータが自動設定される。ここで、感度バランスは、特定音声 V_s の音量が最大となるように、つまり、マイク M1、M2 の音声間の位相差 t_t が閾値 t_t 未満となるように調整される。これにより、特定音声 V_s の収音が継続可能となり、特定音源 S_s を追跡することができる。40

【0064】

例えば、図 10 A に示す例では、発話者 U の会話相手等の特定音源 S_s' が発話者 U の正面に位置し、マイク M1、M2 の音声間の位相差 t_t が閾値 t_t 未満であるので、特定音声 V_s が維持され、不図示の不特定音声 V_n が弱められて入力される。しかし、音源がマイク M2 の側に大きく移動して特定音源 S_s となり、位相差 t_t が閾値 t_t 以上になると、特定音声 V_s が弱められて入力できなくなる。

【0065】

10

20

30

40

50

このため、図 10B に示すように、マイク M1、M2 の信号間のレベルバランスがマイク M2 の側にシフトするように、感度バランスが自動的に調整される。ここで、感度バランスは、マイク M1、M2 と特定音源 Ss の相対的な位置変化に追従して、マイク M1、M2 の音声間の位相差 t が閾値 t 未満となるように調整される。これにより、発話者 U と特定音源 Ss の相対位置が変化しても、特定音声 Vs を連続的に入力することができる。

【0066】

[4-6. 遠隔設定処理]

遠隔設定処理では、受話者 U' による各種パラメータの遠隔設定が可能となる。例えば、受話者 U' は、図 5 に示した設定パネル CP と同様な設定パネル CP' を用いて、各種パラメータを遠隔設定する。

10

【0067】

例えば図 11 に示すように、受話者 U' は、再生装置 100' が発話者 U の送信音声を再生すると、再生音声の品質に応じて、設定パネル CP' 上で各種パラメータを指定（設定）する。再生装置 100' は、受話者 U' の操作に応じて、パラメータ指定情報を通信網 N を介して情報処理装置 100 に送信する。情報処理装置 100 は、パラメータ指定情報に基づいて各種パラメータを設定し、設定状況を設定パネル CP に反映させる。これにより、発話者 U と受話者 U' の間で、パラメータの設定を最適化することで、送信音声の品質をさらに向上させることができる。

20

【0068】

[5. まとめ]

以上説明したように、本実施形態によれば、少なくとも一対で設けられるマイク M1、M2 により収音された外部の音声信号に、少なくともマイク M1、M2 の感度を規定し、少なくともユーザーの指示に応じて設定された処理パラメータに基づいて、ビームフォーミング処理を含む音声処理が施される。これにより、使用環境に応じて、少なくとも収音部の感度を規定する処理パラメータを設定することで、特定音声 Vs が良好な状態で選択的に入力可能となり、送信音声の品質を向上させることができる。

【0069】

以上、添付図面を参照しながら本発明の好適な実施形態について詳細に説明したが、本発明はかかる例に限定されない。本発明の属する技術の分野における通常の知識を有する者であれば、特許請求の範囲に記載された技術的思想の範疇内において、各種の変更例または修正例に想到し得ることは明らかであり、これらについても、当然に本発明の技術的範囲に属するものと了解される。

30

【0070】

例えば、上記実施形態の説明では、使用環境に応じて処理パラメータを設定することで、特定音声 Vs の音声信号のレベルを維持し、不特定音声 Vn の音声信号のレベルを弱める場合について説明した。しかし、特定音声 Vs の音声信号のレベルを弱め、不特定音声 Vn の音声信号のレベルを維持してもよい。これにより、不特定音声 Vn が良好な状態で選択的に入力可能となり、発話者周辺の音声を明瞭に聴取することができる。

【符号の説明】

40

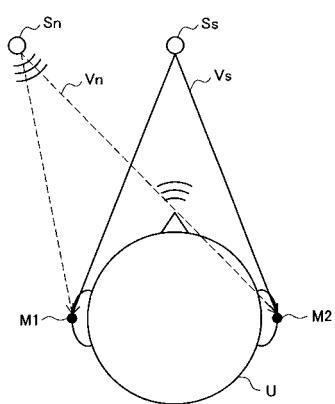
【0071】

- 100 情報処理装置
- 150 音声信号処理部
- 151 感度調整部
- 153 感度調整補正部
- 155 周波数調整部
- 157 時間差分析部
- 159 周波数分析部
- 161 位相差分析部
- 163 ビームフォーミング処理部（BF 処理部）

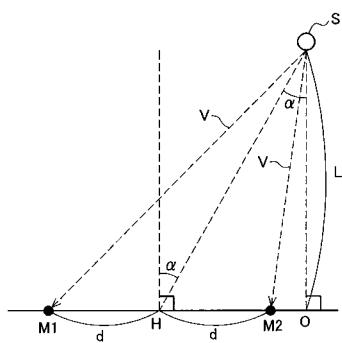
50

U 発話者
 S_s 特定音源
 V_s 特定音声

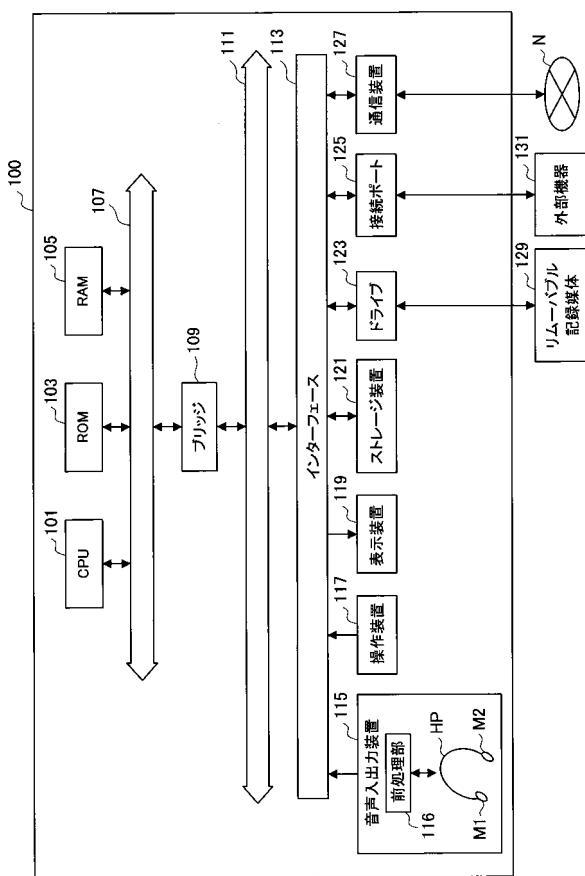
【図 1】



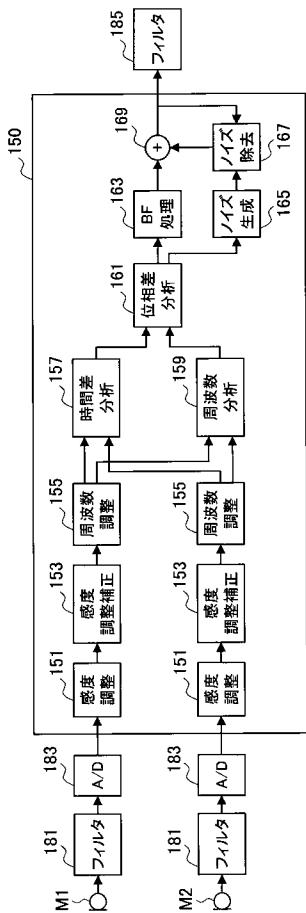
【図 2】



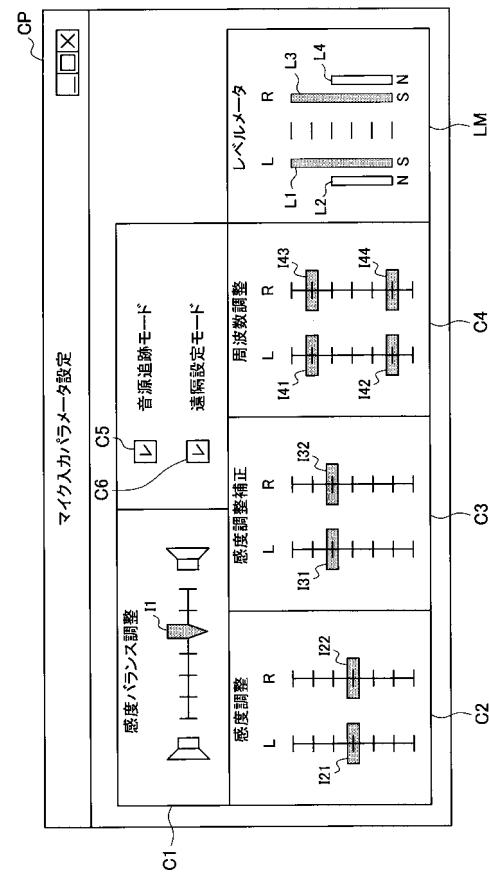
【図 3】



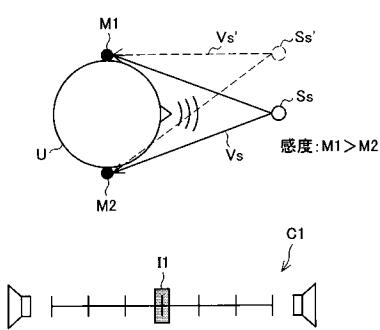
【図4】



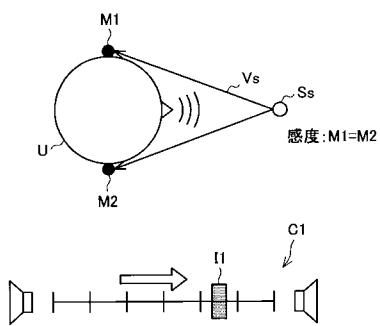
【図5】



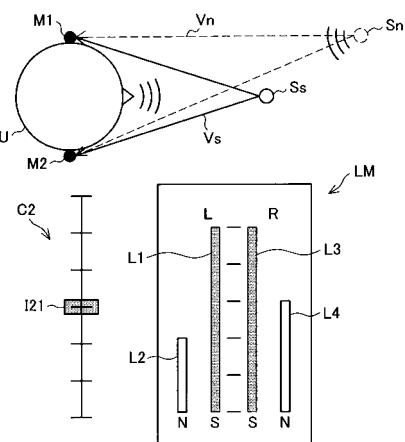
【図6 A】



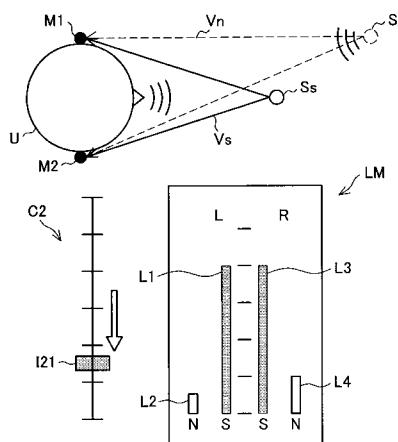
【図6 B】



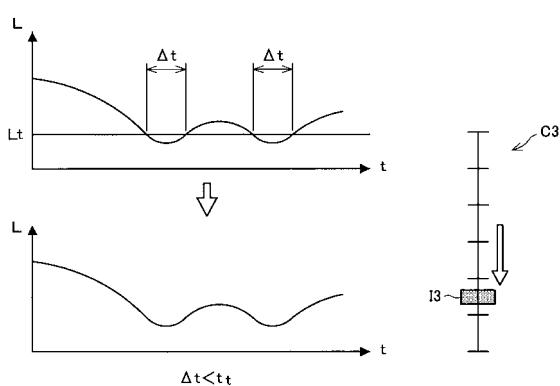
【図7 A】



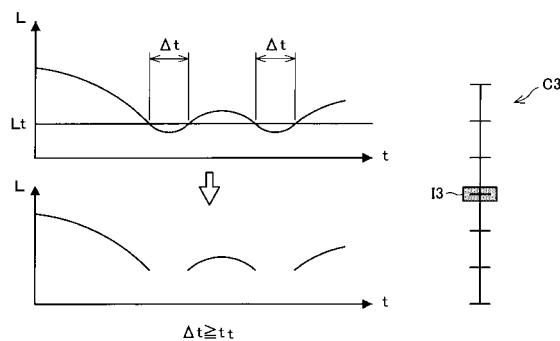
【図 7 B】



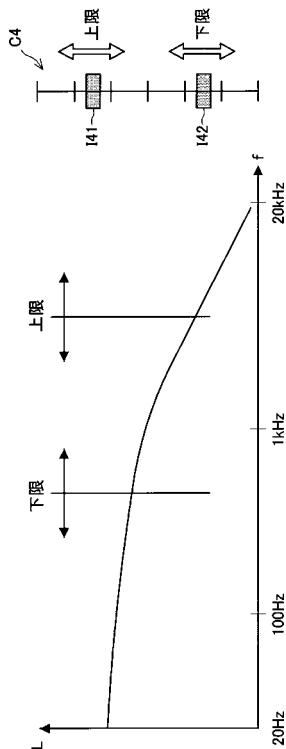
【図 8 B】



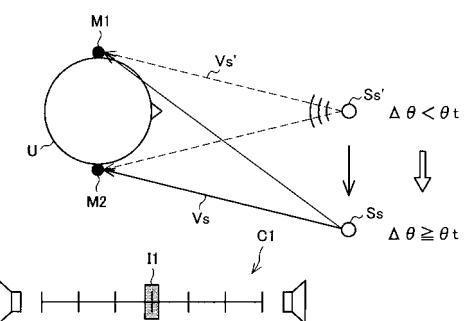
【図 8 A】



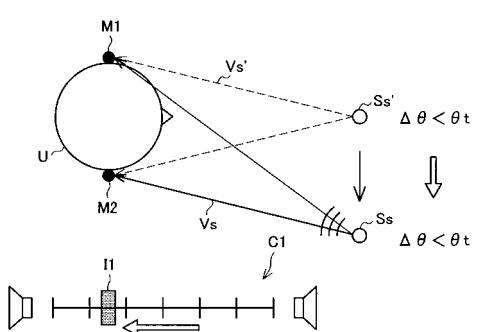
【図 9】



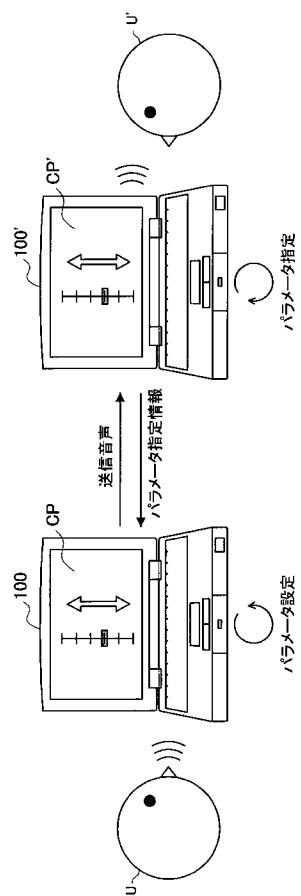
【図 10 A】



【図 10 B】



【図 1 1】



フロントページの続き

(72)発明者 刘 怡君
東京都港区港南1丁目7番1号 ソニー株式会社内

審査官 吉 澤 雅博

(56)参考文献 特開平06-351089(JP,A)
特表2003-527012(JP,A)
特開平09-083988(JP,A)
特開平04-212600(JP,A)

(58)調査した分野(Int.Cl., DB名)

H04R 3/00
H04R 1/40