

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2020-201337

(P2020-201337A)

(43) 公開日 令和2年12月17日(2020.12.17)

(51) Int.Cl.	F I	テーマコード (参考)
G 1 O L 21/034 (2013.01)	G 1 O L 21/034	
G 1 O L 25/84 (2013.01)	G 1 O L 25/84	

審査請求 未請求 請求項の数 18 O L (全 15 頁)

(21) 出願番号	特願2019-106859 (P2019-106859)	(71) 出願人	000004075
(22) 出願日	令和1年6月7日 (2019.6.7)		ヤマハ株式会社
		(74) 代理人	110000970
			特許業務法人 楓国際特許事務所
		(72) 発明者	川合 室登
			静岡県浜松市中区中沢町10番1号 ヤマハ株式会社内

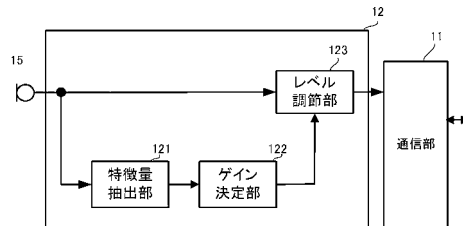
(54) 【発明の名称】 音声処理装置及び音声処理方法

(57) 【要約】

【課題】近端側の話者の音声を適切なレベルで遠端側に送信する音声処理方法を提供する。

【解決手段】音声処理方法は、マイク(15)で收音された音信号から近端側の話者の音声の特徴量を抽出し、抽出された音声の特徴量に基づいて、マイク(15)で收音された音信号のゲインを決定し、決定されたゲインによって処理された信号を遠端側に送信する。

【選択図】図3



**【特許請求の範囲】****【請求項 1】**

マイクと、

前記マイクで収録された音信号から近端側の話者の音声である確率を求める音声推定部と、

前記音声推定部で推定された前記確率に基づいて、前記マイクで収録された前記音信号のゲインを決定するゲイン決定部と、

前記ゲイン決定部で決定された前記ゲインによって処理された信号を遠端側に送信する送信部と、を備えた、音声処理装置。

**【請求項 2】**

前記マイクで収録された前記音信号における音信号対雑音の比率を推定する S N 比算出部をさらに備え、

前記ゲイン決定部は、前記確率と前記 S N 比算出部によって推定された前記比率とに基づいて前記マイクで収録された前記音信号の前記ゲインを決定する、請求項 1 に記載の音声処理装置。

**【請求項 3】**

前記マイクを複数備え、

前記複数のマイクで収録された前記音声の相関を推定する相関算出部をさらに備え、

前記ゲイン決定部は、前記確率及び前記相関を用いて前記複数のマイクで収録された前記音信号の前記ゲインを決定する、請求項 1 又は 2 に記載の音声処理装置。

**【請求項 4】**

前記ゲイン決定部は、前記ゲインを段階的に下げ、又は前記ゲインを瞬時に上げる、請求項 1 乃至 3 のいずれかに記載の音声処理装置。

**【請求項 5】**

前記ゲイン決定部は、前記確率が所定の条件を満たさなかった場合、前記ゲインを最小に決定し、前記確率が前記所定の条件を満たした場合、前記ゲインを最小よりも大きい値に決定する、請求項 1 乃至 4 のいずれかに記載の音声処理装置。

**【請求項 6】**

前記音声推定部は、ケプストラムを用いて、前記確率を求める、請求項 1 乃至 5 のいずれかに記載の音声処理装置。

**【請求項 7】**

前記音声推定部は、機械学習方法を用いて、前記確率を求める、請求項 1 乃至 5 のいずれかに記載の音声処理装置。

**【請求項 8】**

前記音声推定部は、前記マイクで収録された前記音信号の基音及び倍音の周波数分布を用いて、前記確率を求める、請求項 1 乃至 5 のいずれかに記載の音声処理装置。

**【請求項 9】**

マイクと、

前記マイクで収録された音信号から近端側の話者の音声の特徴量を抽出する特徴量抽出部と、

前記特徴量抽出部が抽出した前記音声の特徴量に基づいて、前記マイクで収録された前記音信号のゲインを決定するゲイン決定部と、

前記ゲイン決定部で決定された前記ゲインによって処理された信号を遠端側に送信する送信部と、を備えた、音声処理装置。

**【請求項 10】**

マイクで収録された音信号から近端側の話者の音声である確率を求め、

求められた前記確率に基づいて、前記マイクで収録された前記音信号のゲインを決定し、

決定された前記ゲインによって処理された信号を遠端側に送信する、音声処理方法。

**【請求項 11】**

10

20

30

40

50

前記マイクで収録された前記音信号における音信号対雑音の比率を推定し、  
前記確率と前記比率とに基づいて前記マイクで収録された前記音信号の前記ゲインを決定する、請求項 10 に記載の音声処理方法。

【請求項 12】

複数の前記マイクで収録された前記音声の相関を推定し、  
前記確率及び前記相関を用いて前記マイクで収録された前記音信号の前記ゲインを決定する、請求項 10 又は 11 に記載の音声処理方法。

【請求項 13】

前記ゲインを段階的に下げ、又は前記ゲインを瞬時に上げる、請求項 10 乃至 12 のいずれかに記載の音声処理方法。

【請求項 14】

前記確率が所定の条件を満たさなかった場合、前記ゲインを最小に決定し、前記確率が前記所定の条件を満たした場合、前記ゲインを最小よりも大きい値に決定する、請求項 10 乃至 13 のいずれかに記載の音声処理方法。

【請求項 15】

ケプストラムを用いて、前記確率を求める、請求項 10 乃至 14 のいずれかに記載の音声処理方法。

【請求項 16】

機械学習方法を用いて、前記確率を求める、請求項 10 乃至 14 のいずれかに記載の音声処理方法。

【請求項 17】

前記マイクで収録された前記音信号の基音及び倍音の周波数分布を用いて、前記確率を求める、請求項 10 乃至 14 のいずれかに記載の音声処理方法。

【請求項 18】

マイクで収録された音信号から近端側の話者の音声の特徴量を抽出し、  
抽出した前記音声の特徴量に基づいて、前記マイクで収録された前記音信号のゲインを決定し、  
決定された前記ゲインによって処理された信号を遠端側に送信する、音声処理方法。

【発明の詳細な説明】

【技術分野】

【0001】

この発明の一実施形態は、音信号の信号処理に関する。

【背景技術】

【0002】

従来、マイクから入力された音信号のレベルの強度を測定して、音信号のレベルの強度と予め設定した音声強度閾値とを比較することで、音信号が音声であるか雑音であるかを判定する音声調整装置があった（例えば、特許文献 1 を参照）。この音声調整装置は、所定の音声強度閾値よりも大きい状態（有音状態）で任意の一定時間（5～10 秒）を越えて継続している音信号を、人間の音声ではなく周囲の雑音であるとみなしていた。また、この音声調整装置は、閾値よりも小さい状態（無音状態）が一定（20～30 秒）を超えて継続している音信号に周囲の雑音を含まないとみなしていた。

【先行技術文献】

【特許文献】

【0003】

【特許文献 1】特開 2016 - 85420 号公報

【発明の概要】

【発明が解決しようとする課題】

【0004】

しかしながら、従来の音声調整装置は、該音声調整装置を使用している近端側の話者の音声を、離れた位置（遠端側）にある別の音声調整装置又はスピーカを有する装置に対し

10

20

30

40

50

て、適切なレベルで遠端側に送信することを考慮していなかった。

【0005】

この発明の一実施形態は、近端側の話者の音声を適切なレベルで遠端側に送信することを目的とする。

【課題を解決するための手段】

【0006】

本発明の一実施形態に係る音声処理装置は、マイクと、前記マイクで收音された音信号から近端側の話者の音声を特徴量を抽出する特徴量抽出部と、前記特徴量抽出部が抽出した前記音声を特徴量に基づいて、前記マイクで收音された前記音信号のゲインを決定するゲイン決定部と、前記ゲイン決定部で決定された前記ゲインによって処理された信号を遠端側に送信する送信部と、を備えている。

10

【発明の効果】

【0007】

本発明の一実施形態によれば、近端側の話者の音声を適切なレベルで遠端側に送信することができる。

【図面の簡単な説明】

【0008】

【図1】実施形態1の音声処理システムの構成を示す図である。

【図2】実施形態1の音声処理装置の主要な構成を示すブロック図である。

【図3】実施形態1の音声処理装置の機能的構成を示すブロック図である。

20

【図4】実施形態1のケプストラムの波形の一例であって、図4(A)は、音信号に音声が含まれている波形を示す図であり、図4(B)は、音信号に音声が含まれていない波形を示す図である。

【図5】実施形態1の音声処理装置の動作の一例を示すフローチャートである。

【図6】実施形態1のケプストラムを用いた特徴量の抽出処理の一例を示すフローチャートである。

【図7】実施形態1のゲイン決定部の動作の一例を示すフローチャートである。

【図8】実施形態2の音声処理装置の機能的構成を示すブロック図である。

【図9】実施形態3の音声処理装置の主要な構成を示すブロック図である。

【図10】実施形態3の音声処理装置の機能的構成を示すブロック図である。

30

【図11】音響装置の主要なハードウェア構成を示すブロック図である。

【発明を実施するための形態】

【0009】

(実施形態1)

図1は、実施形態1に係る音声処理システム100の構成を示す図である。音声処理システム100は、ネットワーク2を介して互いに接続される複数の(図1では2つの)音声処理装置1及び音声処理装置3を備えている。

【0010】

音声処理装置1は、第1の地点に設置されている。音声処理装置3は、第2の地点に設置されている。すなわち、音声処理装置1及び音声処理装置3は、互いに遠隔地に設置されている。

40

【0011】

音声処理装置1及び音声処理装置3は、例えば、遠隔地における会議システムに用いられる。音声処理装置1は、例えば、音声処理装置1の利用者の音声を、ネットワーク2を介して、音声処理装置3に送信する。また、音声処理装置3の利用者の音声も、同じく、ネットワーク2を介して、音声処理装置1に送信する。

【0012】

音声処理装置1及び音声処理装置3は、同じ構成及び機能を有する。また、以下の説明では、音声処理装置1の利用者を近端側の利用者又は話者と呼ぶ。また、以下の説明では、音声処理装置3の利用者を遠端側の利用者又は聴者と呼ぶ。さらに、以下の説明では、

50

音声処理装置 1 の話者の音声を、ネットワーク 2 を介して音声処理装置 3 に送信する例で説明する。

【 0 0 1 3 】

なお、以下の実施形態の説明は一例であって、本発明は、以下の例に限定されるものではない。

【 0 0 1 4 】

図 2 は、音声処理装置 1 の主要な構成を示すブロック図である。音声処理装置 1 は、図 2 に示すように、通信部 1 1、CPU 1 2、RAM 1 3、メモリ 1 4、マイク 1 5、スピーカ 1 6、及びユーザインタフェース ( I / F ) 1 7 を備えている。この例でいう、通信部 1 1 は、本発明の送信部に相当する。なお、音声処理装置 3 は、音声処理装置 1 と同じ構成及び同じ機能を有するので、説明を省略する。

10

【 0 0 1 5 】

CPU 1 2 は、記憶媒体であるメモリ 1 4 からプログラムを読み出し、RAM 1 3 に一時記憶することで、種々の動作を行う。

【 0 0 1 6 】

メモリ 1 4 は、フラッシュメモリ又はハードディスクドライブ ( HDD ) 等からなる。メモリ 1 4 は、上記のように CPU 1 2 の動作プログラムを記憶している。メモリ 1 4 は、特徴量抽出プログラム、ゲイン決定プログラム、及び通信制御プログラムを記憶している。特徴量抽出プログラムは、後述の音声の特徴量抽出部 1 2 1 を構成するためのプログラムである。ゲイン決定プログラムは、後述のゲイン決定部 1 2 2 を構成するためのプログラムである。通信制御プログラムは、後述の通信部 1 1 を制御するためのプログラムである。

20

【 0 0 1 7 】

マイク 1 5 は、話者の音声及び周囲の雑音を含む音信号を收音する。マイク 1 5 は、收音した音信号 ( 話者の音声及び周囲の雑音を含む音 ) をデジタル変換する。マイク 1 5 は、デジタル変換した音信号を CPU 1 2 に出力する。

【 0 0 1 8 】

CPU 1 2 は、マイク 1 5 から入力された音信号に信号処理を施す。CPU 1 2 は、信号処理を施した音信号を通信部 1 1 に出力する。なお、音声処理装置 1 は、信号処理の専用プロセッサ ( DSP : Digital Signal Processor ) を備えていてもよい。この場合、CPU 1 2 の指示に従って、DSP が信号処理を行なう。

30

【 0 0 1 9 】

CPU 1 2 は、信号処理として、マイク 1 5 で收音された音信号から、話者の音声である確率 ( 人の声である信頼度 ) を求める。CPU 1 2 は、音声である確率に基づいて、遠端側の聴者にとって近端側の音声をより聞き取りやすくするように音信号に信号処理を施す。なお、信号処理の詳細は後述する。なお、この例でいう、音声である確率とは、0 ~ 1 0 0、又は 0 ~ 1 の値であってもよいし、基準を設けて、該基準と比較した度合いによって求められてもよい。

【 0 0 2 0 】

ユーザ I / F 1 7 は、利用者からの操作を受け付ける。利用者から受け付ける操作は、例えば、スピーカ 1 6 の音量の調整等である。

40

【 0 0 2 1 】

通信部 1 1 は、CPU 1 2 で信号処理された音信号を、ネットワーク 2 を介して、遠端側の音声処理装置 3 に送信する ( 図 1 参照 )。また、通信部 1 1 は、音声処理装置 3 から遠端側の音信号を受信し、スピーカ 1 6 に出力する。

【 0 0 2 2 】

スピーカ 1 6 は、通信部 1 1 が出力した遠端側の音信号 ( 利用者の音声 ) を放音する。また、スピーカ 1 6 は、入力された音信号がデジタル信号である場合、D / A 変換を行ってから音信号を放音する構成であってもよい。

【 0 0 2 3 】

50

C P U 1 2 の詳細な機能及び動作について、図 3 を参照して説明する。図 3 は、音声処理装置 1 の機能的構成を示すブロック図である。C P U 1 2 は、マイク 1 5 で收音された音声を適切なレベルに調節し、通信部 1 1 に出力する。

【 0 0 2 4 】

C P U 1 2 は、図 3 に示すように、特徴量抽出部 1 2 1 と、ゲイン決定部 1 2 2 と、レベル調節部 1 2 3 とを備えている。この例でいう特徴量抽出部 1 2 1 は、本発明の音声推定部に相当する。

【 0 0 2 5 】

特徴量抽出部 1 2 1 は、マイク 1 5 で收音された音信号から、音声である確率として、音声の特徴量を抽出する。より詳細には、特徴量抽出部 1 2 1 は、例えば、ケプストラムを用いて、音声の特徴量を抽出する。

10

【 0 0 2 6 】

図 4 ( A ) は、ケプストラムの波形の一例であって、音信号に音声が含まれている例を示す図である。図 4 ( B ) は、ケプストラムの波形の一例であって、音信号に音声が含まれていない例を示す図である。図 4 ( A ) 及び図 4 ( B ) の縦軸は、レベル ( d B ) を示す。また、図 4 ( A ) 及び図 4 ( B ) の横軸は、ケフレンシー ( m s ) を示す。なお、図 4 ( A ) 及び図 4 ( B ) は、低ケフレンシー成分を拡大したものである。

【 0 0 2 7 】

特徴量抽出部 1 2 1 は、図 4 ( A ) に示すように、例えば、リフタ ( ケプストラムに使用されるフィルタ ) を使用するなどして、一点鎖線で囲まれた低次のケプストラムと二点鎖線で囲まれた高次のケプストラムとに分離する。ここで、低次のケプストラムは、スペクトルの包絡成分に対応する。また、高次のケプストラムはスペクトルの微細成分に対応する。

20

【 0 0 2 8 】

音声は、基音と倍音で構成されている。音声の基音は、高次のケプストラムのピークとして現れる。つまり、特徴量抽出部 1 2 1 は、高次のケプストラムにピークを検出した場合、マイク 1 5 で收音された音信号は、音声である確率が高いと推定する。この例では、特徴量抽出部 1 2 1 は、高次のケプストラムにピークを検出した場合、ピークレベルを 0 ~ 1 の値に変換 ( 正規化 ) し、特徴量としてゲイン決定部 1 2 2 に出力する。

【 0 0 2 9 】

一方、特徴量抽出部 1 2 1 は、図 4 ( B ) に示すように、高次のケプストラムにピークを検出しない場合、音声である確率が低いと推定する。特徴量抽出部 1 2 1 は、高次のケプストラムにピークが無い旨 ( 特徴量を 0 とする旨 ) をゲイン決定部 1 2 2 に出力する。

30

【 0 0 3 0 】

ゲイン決定部 1 2 2 は、特徴量抽出部 1 2 1 によって抽出された特徴量 ( 0 ~ 1 の値 ) に基づいて、マイク 1 5 で收音した音信号のゲインを決定する。ゲイン決定部 1 2 2 は、例えば、特徴量が 0 ~ 1 の場合、ゲインを 0 ( 最小値 ) よりも大きな値に決定する。また、ゲイン決定部 1 2 2 は、抽出された特徴量が 0 の場合、ゲインを 0 ( 最小値 ) に決定する。

【 0 0 3 1 】

なお、ゲイン決定部 1 2 2 は、受け取った特徴量に応じて、ゲインを 0 ~ 1 の間に決定してもよい。また、ゲイン決定部 1 2 2 は、予め閾値を決めておき、受け取った特徴量と閾値とを比較して、ゲインを決定してもよい。この場合、例えば、ゲイン決定部 1 2 2 は、特徴量が閾値以下であれば、ゲインを 0 に決定する。また、この場合、例えば、特徴量が閾値を越えていれば、ゲインを 0 ~ 1 に決定する。

40

【 0 0 3 2 】

レベル調節部 1 2 3 は、ゲイン決定部 1 2 2 によって決定されたゲインを用いて、マイク 1 5 で收音した音信号のレベルを調節する。

【 0 0 3 3 】

音声処理装置 1 は、マイク 1 5 で收音された音信号が音声である確率 ( 特徴量 ) が高け

50

れば、ゲインを1又は確率に応じて0～1の間に決定する。また、マイク15で收音された音信号が音声である確率が低ければ、ゲインを0に決定する。これにより、音声処理装置1は、近端側の不要な雑音を抑制することができる。従って、遠端側の聴者は、適切なレベルで近端側の話者の音声を聞くことができる。

#### 【0034】

音声処理装置1の動作について、図5を参照して説明する。図5は、音声処理装置1の動作の一例を示すフローチャートである。

#### 【0035】

特徴量抽出部121は、マイク15で收音された音信号から、音声の確率として、音声の特徴量を抽出する(S1)。ゲイン決定部122は、音声の特徴量に基づいて、マイク15で收音された音信号のゲインを決定する(S2)。レベル調節部123は、決定されたゲインを用いて音信号のレベルを調節する(S3)。送信部は、処理された信号を遠端側に送信する(S4)。

10

#### 【0036】

次に、特徴量抽出部121による特徴量の抽出処理の一例について、図6を参照して説明する。図6は、ケプストラムを用いた特徴量の抽出処理の一例を示すフローチャートである。

#### 【0037】

まず、特徴量抽出部121は、マイク15で收音された音信号にFFT(Fast Furiers Transform)(高速フーリエ変換処理)を施し、FFTを施した信号の絶対値を計算する。これにより、特徴量抽出部121は、音信号の振幅スペクトルを算出する(S11)。次に、特徴量抽出部121は、振幅スペクトルを対数計算(Log計算)し、振幅スペクトルをデシベル(dB)に変換する。これにより、特徴量抽出部121は、対数スペクトルを算出する(S12)。さらに、特徴量抽出部121は、対数スペクトルにDCT(Discrete Cosine Transform)(離散コサイン変換処理)を施す。これにより特徴量抽出部121は、ケプストラム(波形)を算出する(S13)。特徴量抽出部121は、算出したケプストラムの波形から高次のケプストラムのピークレベルを検出する(S14)。特徴量抽出部121は、検出したピークレベルを正規化した特徴量をゲイン決定部122に出力する(S15)。

20

#### 【0038】

なお、特徴量抽出部121は、高次ケプストラムにピークレベルを検出しなかった場合、音声の特徴量(音声である確率)を0としてゲイン決定部122に出力する。

30

#### 【0039】

次に、ゲイン決定部122の動作の一例について、図7を参照して説明する。図7は、ゲイン決定部122の動作の一例を示すフローチャートである。なお、以下の説明は一例であって、限定されるものではない。

#### 【0040】

ゲイン決定部122は、特徴量抽出部121から特徴量を受け取る(S21)。ゲイン決定部122は、受け取った特徴量に応じて、ゲインを決定する(S22)。ゲイン決定部122は、受け取った特徴量が0～1であれば、又は予め決められた閾値を越えていれば、(S22:Yes)、ゲインを0よりも大きい値に決定し、レベル調節部123に出力する(S23)。

40

#### 【0041】

一方、ゲイン決定部122は、特徴量が0又は特徴量が閾値よりも低ければ、(S22:No)、ゲインを0に決定し、レベル調節部123に出力する(S23)。

#### 【0042】

実施形態1の音声処理装置1は、ケプストラムの波形のピークレベルから求められた音声である確率(音声の特徴量)に基づいて、ゲインを決定する。音声処理装置1は、音声である確率(音声の特徴量)に基づいて決定されたゲインで、マイク15で收音された音信号のレベルを調節する。これにより、実施形態1の音声処理装置1は、近端側の話者の

50

音声を適切なレベルで遠端側に送信することができる。

【 0 0 4 3 】

なお、ゲイン決定部 1 2 2 は、例えば、受け取った特徴量が 1 又は閾値を越えた場合、瞬時にゲインを上げてよい。この場合、音声処理装置 1 は、音声の始まり部分を聞き取りやすくする（音声の頭抜けを抑制することができる）。また、ゲイン決定部 1 2 2 は、受け取った特徴量が 0 又は閾値以下の場合、ゲインを時間の経過とともに段階的に下げてもよい。この場合、音声処理装置 1 は、急激に音量が下がらないので、利用者に違和感を覚えさせない。

【 0 0 4 4 】

（実施形態 2）

実施形態 2 の音声処理装置 1 及び音声処理装置 3 の CPU 1 2 A について、図 8 を参照して説明する。図 8 は、実施形態 2 の CPU 1 2 A の機能的構成を示すブロック図である。なお、上述の実施形態 1 の CPU 1 2 と同じ構成には、同じ符号を付し、説明を省略する。また、実施形態 2 の音声処理装置 1 及び 3 は同じ構成及び同じ機能を有するものとし、以下の説明では、音声処理装置 1 の CPU 1 2 A を代表して説明する。

【 0 0 4 5 】

実施形態 2 の CPU 1 2 A は、図 8 に示すように、マイク 1 5 で收音された音信号に対する、周囲の雑音の SN (Signal to Noise) 比率を算出する SN 比算出部 1 2 4 を備えている。すなわち、CPU 1 2 A は、SN 比率の算出を、プログラムを読み出して実行する。

【 0 0 4 6 】

SN 比算出部 1 2 4 は、音信号と、雑音（例えば、タイピング音、空調の音、装置から遠い位置の音など、話者の音声以外の音）との SN 比率を算出する。SN 比率は、例えば、以下の数式 1 よって算出される。

【 0 0 4 7 】

【数 1】

$$\text{SN 比率} = \frac{P_S}{P_N}$$

【 0 0 4 8 】

上記の数式 1 において、 $P_S$  は、マイク 1 5 で收音された音信号のパワーである。また、 $P_N$  は、マイク 1 5 で收音された雑音のパワーである。この場合、SN 比算出部 1 2 4 は、例えば、予め、近端側の話者が音声を発していない（話していない）状態において、雑音のパワーを算出しておく。SN 比算出部 1 2 4 は、予め算出していた雑音のパワーと、收音された音信号のパワーとを用いて SN 比率を算出する。SN 比率が大きい場合、音声に対して雑音が小さいことを示す。また、SN 比率が小さい場合、音声に対して雑音が大きいことを示す。

【 0 0 4 9 】

ゲイン決定部 1 2 2 は、例えば、SN 比算出部 1 2 4 によって算出された SN 比率と特徴量抽出部 1 2 1 によって抽出された特徴量との積に応じてゲインを決定する。

【 0 0 5 0 】

なお、ゲイン決定部 1 2 2 は、予め決められた閾値と、上記積とを比較してもよい。この場合、ゲイン決定部 1 2 2 は、上記積が閾値を越えていれば、ゲインを 0 ~ 1 に決定する。また、ゲイン決定部 1 2 2 は、上記積が閾値以下であれば、ゲインを 0 に決定する。

【 0 0 5 1 】

また、ゲイン決定部 1 2 2 は、特徴量抽出部 1 2 1 によって抽出された音声の特徴量と、SN 比算出部 1 2 4 によって算出された SN 比率特徴量抽出部との積に応じてゲインを決定する例に限定されない。ゲイン決定部 1 2 2 は、例えば、音声の特徴量及び算出した SN 比率を論理式、例えば、論理積、論理和、排他的論理和等、又は複数の論理式の組み合わせを用いてゲインを決定してもよい。

10

20

30

40

50

## 【 0 0 5 2 】

実施形態 2 の音声処理装置 1 は、近端側の話者の特徴量抽出部 1 2 1 による音声である確率（音声の特徴量）だけでなく、音信号対雑音の S N 比率にも基づいて、音信号のレベルを調節する。これにより、音声処理装置 1 は、マイク 1 5 で收音された近端側の話者の音声をより適切なレベルで遠端側に送信することができる。

## 【 0 0 5 3 】

なお、実施形態 2 の音声処理装置 1 では、C P U 1 2 A が S N 比算出プログラムを読みだして実行している例で説明したが、S N 比算出部は、S N 比算出回路で構成されていてもよい。

## 【 0 0 5 4 】

（実施形態 3）

実施形態 3 の音声処理装置 1 A について図 9 を参照して説明する。図 9 は、実施形態 3 の音声処理装置 1 A の主要な構成を示すブロック図である。図 1 0 は、実施形態 3 の音声処理装置 1 A の C P U 1 2 B の機能的構成を示すブロック図である。なお、音声処理装置 3 は、音声処理装置 1 A と同じ構成及び機能を有しているので、以下の説明では、音声処理装置 1 A を代表して説明する。また、音声処理装置 1 と同じ構成には、同じ符号を付し、説明を省略する。

## 【 0 0 5 5 】

実施形態 3 の音声処理装置 1 A は、図 9 に示すように、複数（図 9 では 2 つ）のマイク 1 5 A 及びマイク 1 5 B を備えている。音声処理装置 1 A は、マイク 1 5 A 及びマイク 1 5 B のそれぞれから近端側の音信号を收音する。

## 【 0 0 5 6 】

マイク 1 5 A 及びマイク 1 5 B のそれぞれは、音信号を C P U 1 2 B に出力する。

## 【 0 0 5 7 】

C P U 1 2 B は、図 1 0 に示すように、相関算出部 1 2 5 を備えている。相関算出部 1 2 5 は複数のマイク 1 5 A 及びマイク 1 5 B で收音した音信号の相関を算出する。相関算出部 1 2 5 は、白色化相互関数を算出する。

## 【 0 0 5 8 】

相関算出部 1 2 5 は、マイク 1 5 A から生成される第 1 音信号  $S_1(t)$  のスペクトル及びマイク 1 5 B から生成される第 2 音信号  $S_2(t)$  のスペクトルの相関を求める。

## 【 0 0 5 9 】

なお、以下の説明では、一例として、マイク 1 5 A は、指向性マイクである。マイク 1 5 B は、全方向に均一な感度を有する無指向性マイクである。ただし、マイク 1 5 A 及びマイク 1 5 B の指向性の態様は、この例に限るものではない。例えば、マイク 1 5 A 及びマイク 1 5 B とともに無指向性のマイクであってもよいし、ともに指向性のマイクであってもよい。また、マイクの数も 2 つに限るものではなく、音声処理装置 1 A は、例えば 3 つ以上のマイクを備えていてもよい。

## 【 0 0 6 0 】

白色化相互関数は、例えば、以下の数式 2 ~ 数式 6 によって算出される。まず、相関算出部 1 2 5 は、以下の数式 2 及び数式 3 に従って、第 1 音信号  $S_1(t)$  及び第 2 音信号  $S_2(t)$  のそれぞれに F F T を施す。そして、相関算出部 1 2 5 は、第 1 音信号  $S_1(t)$  及び第 2 音信号  $S_2(t)$  を周波数領域の信号  $S_1(\omega)$  及び信号  $S_2(\omega)$  に変換する。

## 【 0 0 6 1 】

【数 2】

$$S_1(\omega) = \int_{-\infty}^{\infty} s_1(t) e^{-j\omega t} dt$$

## 【 0 0 6 2 】

10

20

30

40

【数 3】

$$S_2(\omega) = \int_{-\infty}^{\infty} s_2(t) e^{-j\omega t} dt$$

【0063】

次に、相関算出部 125 は、周波数領域の信号  $S_1(\omega)$  及び信号  $S_2(\omega)$  の相互相関関数  $S_{1,2}(\omega)$  を算出する。

【0064】

【数 4】

$$S_{1,2}(\omega) = \frac{S_1(\omega) S_2(\omega)^*}{|S_1(\omega)| |S_2(\omega)|}$$

10

【0065】

上記の数式 4 における、\* は、複素共役（複素数について、虚部を -1 倍する）を示す。

【0066】

さらに、相関算出部 125 は、数式 4 で算出された相互相関関数  $S_{1,2}(\omega)$  から、例えば、IFFT (Inverse Fast Fourier Transfer (フーリエ逆変換)) を施して白色化相互相関関数係数  $CPS_{1,2}(\tau)$  を算出する。

【0067】

【数 5】

$$CPS_{1,2}(\tau) = \int_{-\infty}^{\infty} S_{1,2}(\omega) e^{-j\omega\tau} d\omega$$

20

【0068】

さらに白色化相互相関関数係数  $CPS_{1,2}(\tau)$  から、白色化相互相関関数係数が最大となる時間差  $TDOA_{1,2}$  を算出する。

【0069】

【数 6】

$$TDOA_{1,2} = \operatorname{argmax}(CSP_{1,2}(\tau))$$

30

【0070】

相関算出部 125 は、上述の数式 2 ~ 数式 6 を用いることで、第 1 音信号及び第 2 音信号  $S_2(t)$  の波形の時間差（位相差） $TDOA_{1,2}$  を求める。時間差  $TDOA_{1,2}$  が大きい場合、相関算出部 125 は、装置の近くに音源が存在するとみなし、音源からの音声を近端側の話者の音声であると判定する。一方、相関算出部 125 は、時間差  $TDOA_{1,2}$  が小さい場合、間接音が多く、装置の遠方の音源からの雑音であると判定する。

【0071】

ゲイン決定部 122 は、特徴量抽出部 121 によって抽出された音声の特徴量と相関算出部 125 によって判定された判定結果の積を用いて収音した音信号のゲインを決定する。

40

【0072】

実施形態 3 の音声処理装置 1A の相関算出部 125 は、白色化相互相関関数を用いることで、音源からの音信号が音声又は遠方の雑音であるかを判定する。これにより、実施形態 3 の音声処理装置 1A は、近端側の話者の音声である確率（音声の特徴量）だけでなく、遠方からの雑音であるか否かにより、音信号のレベルを調節する。従って、音声処理装置 1A は、マイク 15 で収音された近端側の話者の音声をより適切なレベルで遠端側に送信することができる。

【0073】

なお、ゲイン決定部 122 は、音声である確率（音声の特徴量）及び相関算出部 125 による判定結果の積を用いてゲインを決定する例に限定されない。ゲイン決定部 122 は

50

、例えば、音声である確率（音声の特徴量）及び該判定結果を論理式、例えば、論理積、論理和、排他的論理和等、又は複数の論理式の組み合わせでゲインを決定してもよい。

【0074】

また、実施形態3の音声処理装置1Aは、実施形態2の音声処理装置1と組み合わせて使用してもよい。この場合、例えば、ゲイン決定部122は、音声である確率+SN比率×該判定結果により、マイク15で収録された音信号のゲインを決定する。これにより、音声処理装置1Aは、より精度よく、マイク15で収録された近端側の話者の音声を適切なレベルで遠端側に送信することができる。なお、ゲイン決定部122は、音声である確率（音声の特徴量）、SN比率及び該判定結果を、複数の論理式、例えば論理積、論理和、排他的論理和等、を組み合わせることでゲインを決定してもよい。

10

【0075】

また、実施形態3の音声処理装置1Aでは、CPU12Bが相関算出プログラムを読みだして実行している例で説明したが、相関算出125は、相関算出回路で構成されていてもよい。

【0076】

（変形例1）

変形例1の音声処理装置1の特徴量抽出部121について、説明する。音声処理装置1は、機械学習によって、音信号が近端側の話者の音声であるかどうか判定する。

【0077】

音声処理装置1は、例えば、予め複数の音声サンプルを記憶させておく。音声処理装置1は、複数の音声サンプルの特徴量を算出し、音声の特徴量を学習モデルとしてパターン化しておく。音声処理装置1は、マイク15で収録した音信号を、記憶されていた音声サンプルの特徴量と比較する。音声処理装置1は、マイク15で収録した音声の特徴量と、記憶されていた音声サンプルの特徴量のパターンマッチングにより、音声である確率を求める。

20

【0078】

例えば、音声処理装置1は、音声の特徴量及び、雑音の特徴量を学習し、それぞれの音声モデル及び、雑音モデルとしてパターン化しておく。音声処理装置1は、未知の（マイクで収録した）音信号を入力し、より尤度の高い値を示すモデルを基に、該音信号が音声か否かが判定する。

30

【0079】

変形例1の音声処理装置1は、機械学習により、マイク15で収録した音信号から音声の確率を求める。これにより、変形例1の音声処理装置1は、より簡易な方法で、マイク15で収録された近端側の話者の音声を適切なレベルで遠端側に送信することができる。

【0080】

なお、変形例1の音声処理装置1は、ニューラルネットワークを用いて、音声である確率を求めてもよい。すなわち、変形例1の音声処理装置1は、ニューラルネットワークによって予め学習された、未知の音信号の特徴量を入力し、出力結果に基づいて、音信号が音声か否かを判定する。

【0081】

（変形例2）

変形例2の音声処理装置1の特徴量抽出部121について、説明する。変形例2の音声処理装置1の特徴量抽出部121は、音信号の基音と倍音の周波数分布によって、音声の特徴量を抽出する。

【0082】

特徴量抽出部121は、例えば、マイク15で収録された音信号にFFTを施す。特徴量抽出部121は、マイク15で収録された音信号から基音と基音の整数倍である倍音を検出すれば、話者の音声であると推定する。

【0083】

変形例2の音声処理装置1は、より簡易な方法で音声である確率を求める（音声の特徴

50

量を抽出する)ことができる。

【0084】

(別の変形例)

別の変形例について、以下に列挙する。

【0085】

別の変形例の音声処理装置1の特徴量抽出部121は、周波数帯域におけるパワー(dB)を算出する。より詳細には、特徴量抽出部121は、音信号にFFTを施して、パワーを算出する。特徴量抽出部121は、マイク15で収録した音信号のパワーに応じて、音信号が音声である確率を求める。言い換えると、特徴量抽出部121は、マイク15で収録した音信号のパワーに応じて、音信号が音声であるか否かを判定する。

10

【0086】

また、別の変形例の音声処理装置1の特徴量抽出部121は、スペクトルフラックス手法によって音声の特徴量を抽出する。スペクトルフラックスは、音声スペクトルの局所的な変化を示すものである。すなわち、スペクトルフラックスが大きければ、音の変動が大きいことを示す。すなわち、音声処理装置1の特徴量抽出部121は、スペクトルフラックスが大きければ、近端側の話者が発話したと判断する。

【0087】

また、別の変形例の音声処理装置1の特徴量抽出部121は、非周期性指標によって音声の特徴量を抽出する。非周期性指標は、音声の周期成分及び非周期成分の割合を示す。非周期性指標は、囁れ声の子音のように周期的な成分と非周期的な成分が混在した音を表すために用いられている。本変形例の特徴量抽出部121は、非周期性指標を算出する。すなわち、音声処理装置1の特徴量抽出部121は、非周期性指標を検出すれば、近端側の話者が発話していると判定する。

20

【0088】

また、別の変形例の特徴量抽出部121は、音声である確率(音声の特徴量)として、音信号の時間波形がゼロと交差する回数を計算する。特徴量抽出部121は、音声の時間領域においてゼロ交差となる点の出現頻度を算出したものである。ゼロ交差は、音声の基音であるピッチに対応する。すなわち、音声処理装置1の特徴量抽出部121は、ゼロ交差を検出すれば、近端側の話者の音声の確率が高いと推定する。

【0089】

なお、本発明の構成は、全て、FPGA(Field-Programmable Gate Array)などのハードウェアにより実現してもよい。例えば、図11に示すように、音声処理装置1又は音声処理装置1Aは、特徴量抽出部121は、特徴量抽出部121に対応する特徴量抽出回路121A、ゲイン決定部122に対応するゲイン決定回路122Aを備えていてもよい。

30

【0090】

また、通信部11のハードウェアの機能を、CPU12、CPU12A又はCPU12Bがメモリ14に記憶された通信制御プログラムを読み出して実行することで実現してもよい。

【0091】

また、本発明の構成は、適宜ハードウェア及びソフトウェアを組み合わせることで実現してもよい。

40

【0092】

上述の実施形態及び変形例の説明は、すべての点で例示であって、制限的なものではないと考えられるべきである。本発明の範囲は、上述の実施形態及び変形例ではなく、特許請求の範囲によって示される。さらに、本発明の範囲には、特許請求の範囲と均等の意味及び範囲内でのすべての変更が含まれることが意図される。

【符号の説明】

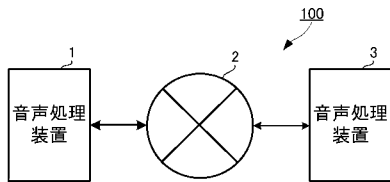
【0093】

1, 1A, 3... 音声処理装置

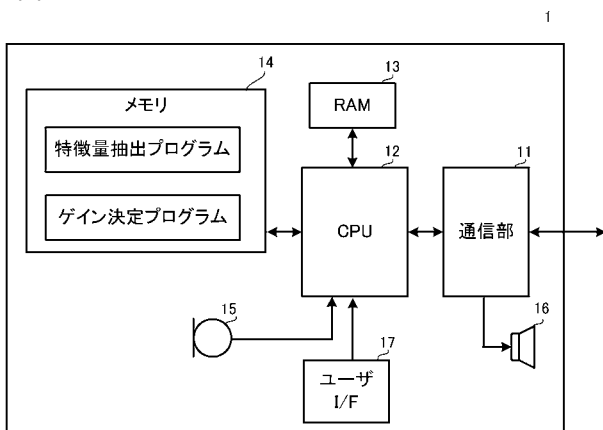
50

- 1 1 ... 通信部 (送信部)
- 1 5 ... マイク
- 1 2 1 ... 特徴量抽出部 (音声推定部)
- 1 2 2 ... ゲイン決定部
- 1 2 4 ... S N 比算出部
- 1 2 5 ... 相関算出部

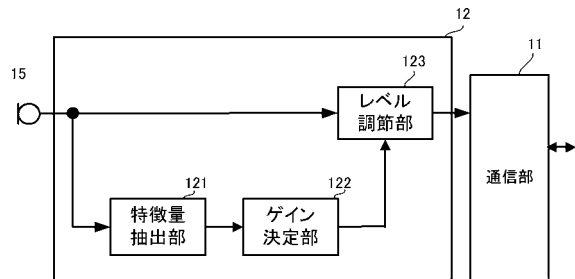
【 図 1 】



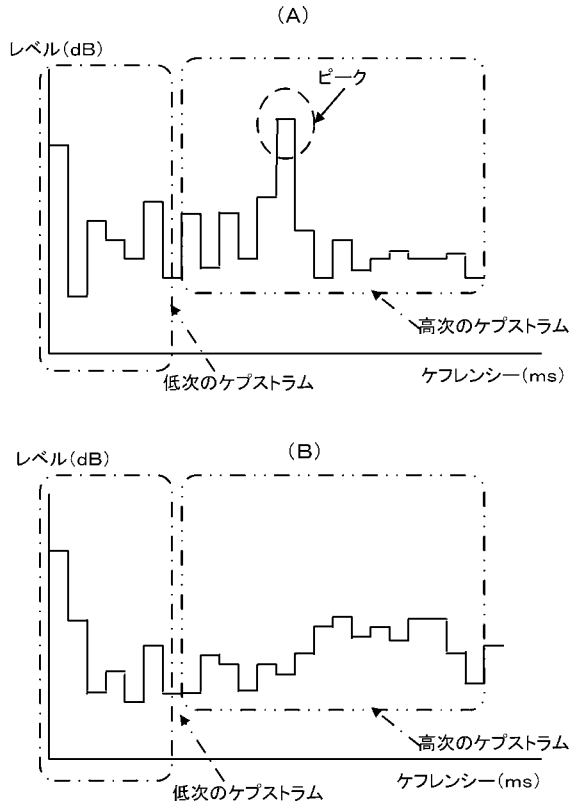
【 図 2 】



【 図 3 】



【 図 4 】



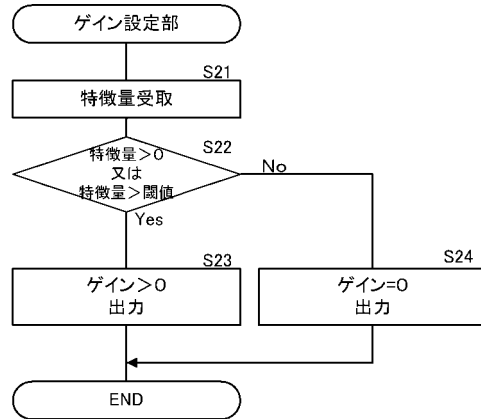
【 図 5 】



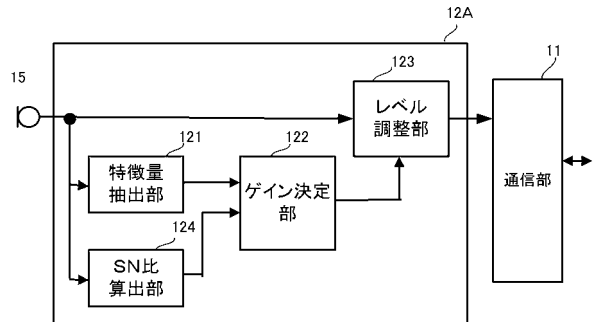
【 図 6 】



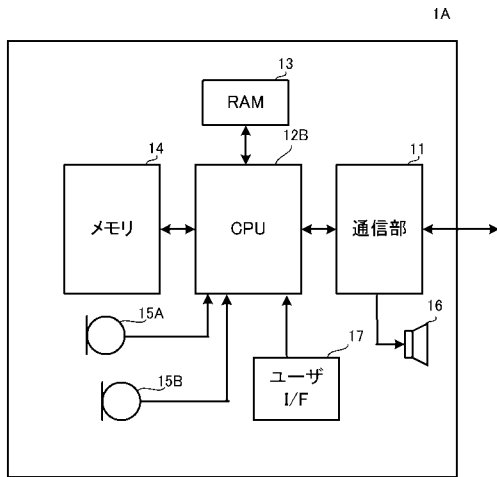
【 図 7 】



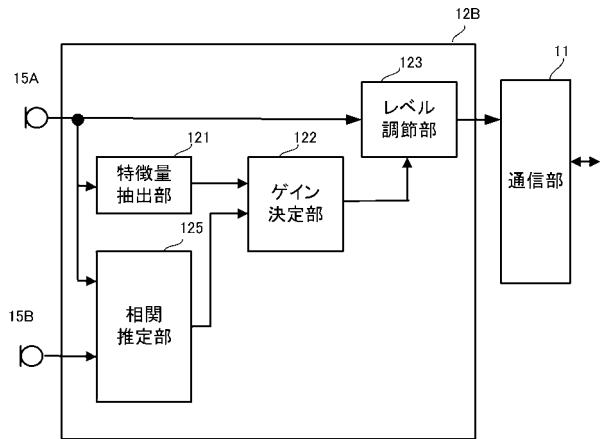
【 図 8 】



【図9】



【図10】



【図11】

