

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6794887号
(P6794887)

(45) 発行日 令和2年12月2日(2020.12.2)

(24) 登録日 令和2年11月16日(2020.11.16)

(51) Int.Cl.		F I	
G 1 O L 21/0232 (2013.01)		G 1 O L 21/0232	
G 1 O L 21/028 (2013.01)		G 1 O L 21/028	B
G 1 O L 21/034 (2013.01)		G 1 O L 21/034	
G 1 O L 25/51 (2013.01)		G 1 O L 25/51	4 0 0
G 1 O L 25/78 (2013.01)		G 1 O L 25/78	

請求項の数 7 (全 22 頁)

(21) 出願番号 特願2017-54257 (P2017-54257)
 (22) 出願日 平成29年3月21日 (2017. 3. 21)
 (65) 公開番号 特開2018-155996 (P2018-155996A)
 (43) 公開日 平成30年10月4日 (2018. 10. 4)
 審査請求日 令和1年12月12日 (2019. 12. 12)

(73) 特許権者 000005223
 富士通株式会社
 神奈川県川崎市中原区上小田中4丁目1番
 1号
 (74) 代理人 100099759
 弁理士 青木 篤
 (74) 代理人 100119987
 弁理士 伊坪 公一
 (74) 代理人 100133835
 弁理士 河野 努
 (74) 代理人 100135976
 弁理士 宮本 哲夫
 (72) 発明者 松尾 直司
 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

最終頁に続く

(54) 【発明の名称】 音声処理用コンピュータプログラム、音声処理装置及び音声処理方法

(57) 【特許請求の範囲】

【請求項1】

第1の音声入力部により生成された第1の音声信号、及び、前記第1の音声入力部と異なる位置に配置された第2の音声入力部により生成された第2の音声信号を、それぞれ、所定の時間長を持つフレームごとに周波数領域の第1の周波数スペクトル及び第2の周波数スペクトルに変換し、

フレームごとに、前記第1の周波数スペクトル及び前記第2の周波数スペクトルに基づいて、受音することが優先される第1の方向及び前記第1の方向と異なる第2の方向のうちの前記第2の方向に位置する音源のみが音声を発した確からしさを算出し、

フレームごとに、前記第1の周波数スペクトル及び前記第2の周波数スペクトルに基づいて算出される前記第1の方向から到来する音声を含む第1の指向音声信号を出力するとともに、前記確からしさに応じて、前記第1の周波数スペクトル及び前記第2の周波数スペクトルに基づいて算出される前記第2の方向から到来する音声を含む第2の指向音声信号を出力するか否かを制御する、

ことをコンピュータに実行させるための音声処理用コンピュータプログラム。

【請求項2】

前記第2の指向音声信号の出力を制御することは、前記確からしさが第1の閾値よりも高くなるフレームについて前記第2の指向音声信号を出力する、請求項1に記載の音声処理用コンピュータプログラム。

【請求項3】

10

20

前記第2の指向音声信号の出力を制御することは、第1のフレームにおける前記確からしさが前記第1の閾値よりも低い第2の閾値未満となり、かつ、前記第1のフレームの直前のフレームにおける前記確からしさが前記第2の閾値以上である場合、前記第1のフレームから第1の期間経過後のフレームから前記第2の指向音声信号の出力を停止する、請求項2に記載の音声処理用コンピュータプログラム。

【請求項4】

フレームごとに、前記第1の周波数スペクトル及び前記第2の周波数スペクトルに基づいて、前記第1の指向音声信号のパワー及び前記第2の指向音声信号のパワーを算出することをさらにコンピュータに実行させ、

前記確からしさを算出することは、フレームごとに、前記第1の指向音声信号のパワーに対する前記第2の指向音声信号のパワーの比率に基づいて前記確からしさを算出する、請求項1～3の何れか一項に記載の音声処理用コンピュータプログラム。

10

【請求項5】

フレームごとに、前記第1の周波数スペクトル及び前記第2の周波数スペクトルに基づいて、前記第1の指向音声の非定常性度合い及び前記第2の指向音声の非定常性度合いを算出することをさらにコンピュータに実行させ、

前記確からしさを算出することは、フレームごとに、前記第1の指向音声の非定常性度合いに対する前記第2の指向音声の非定常性度合いの非定常度比と前記パワー比の和に基づいて前記確からしさを算出する、請求項4に記載の音声処理用コンピュータプログラム。

20

【請求項6】

集音した音声を表す第1の音声信号を生成する第1の音声入力部と、

前記第1の音声入力部と異なる位置に配置され、集音した音声を表す第2の音声信号を生成する第2の音声入力部と、

前記第1の音声信号及び第2の音声信号を、それぞれ、所定の時間長を持つフレームごとに周波数領域の第1の周波数スペクトル及び第2の周波数スペクトルに変換する時間周波数変換部と、

フレームごとに、前記第1の周波数スペクトル及び前記第2の周波数スペクトルに基づいて、受音することが優先される第1の方向及び前記第1の方向と異なる第2の方向のうちの前記第2の方向に位置する音源のみが音声を発した確からしさを算出する音源方向判定部と、

30

フレームごとに、前記第1の周波数スペクトル及び前記第2の周波数スペクトルに基づいて算出される前記第1の方向から到来する音声を含む第1の指向音声信号を出力するとともに、前記確からしさに応じて、前記第1の周波数スペクトル及び前記第2の周波数スペクトルに基づいて算出される前記第2の方向から到来する音声を含む第2の指向音声信号を出力するか否かを制御する指向特性制御部と、

を有する音声処理装置。

【請求項7】

第1の音声入力部により生成された第1の音声信号、及び、前記第1の音声入力部と異なる位置に配置された第2の音声入力部により生成された第2の音声信号を、それぞれ、所定の時間長を持つフレームごとに周波数領域の第1の周波数スペクトル及び第2の周波数スペクトルに変換し、

40

フレームごとに、前記第1の周波数スペクトル及び前記第2の周波数スペクトルに基づいて、受音することが優先される第1の方向及び前記第1の方向と異なる第2の方向のうちの前記第2の方向に位置する音源のみが音声を発した確からしさを算出し、

フレームごとに、前記第1の周波数スペクトル及び前記第2の周波数スペクトルに基づいて算出される前記第1の方向から到来する音声を含む第1の指向音声信号を出力するとともに、前記確からしさに応じて、前記第1の周波数スペクトル及び前記第2の周波数スペクトルに基づいて算出される前記第2の方向から到来する音声を含む第2の指向音声信号を出力するか否かを制御する、

50

ことを含む音声処理方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、例えば、複数のマイクロホンを用いて集音された音声を含む音声信号を処理する音声処理用コンピュータプログラム、音声処理装置及び音声処理方法に関する。

【背景技術】

【0002】

近年、複数のマイクロホンにより音声を集音することで得られた音声信号を処理する音声処理装置が開発されている。このような音声処理装置において、音声信号に含まれる特定方向からの音声を聞き取り易くするために、その音声信号においてその特定方向以外からの音声を抑圧する技術が研究されている（例えば、特許文献1及び2を参照）。

10

【先行技術文献】

【特許文献】

【0003】

【特許文献1】特開2007-318528号公報

【特許文献2】特開2011-139378号公報

【発明の概要】

【発明が解決しようとする課題】

【0004】

20

場合によっては、特定方向に位置する音源からの音声だけでなく、他の方向に位置する他の音源からの音声についても、抑圧しないことが好ましいことがある。しかし、例えば、特許文献1に記載された技術では、特定方向以外の方向から到来する音声は抑圧されてしまう。一方、例えば、特許文献2に記載された技術では、特定方向に位置する音源からの音声だけでなく、想定される他の方向に位置する他の音源からの音声も抑圧しないようにすると、抑圧対象とならない方向の範囲が広過ぎて、雑音の抑圧が不十分となる。その結果として、特定方向に位置する音源からの音声の聞き取り易さが十分に向上しない可能性がある。

【0005】

一つの側面では、本発明は、優先する方向に位置する音源からの音声だけでなく、他の方向に位置する他の音源からの音声も抑圧せずに出力できる音声処理用コンピュータプログラムを提供することを目的とする。

30

【課題を解決するための手段】

【0006】

一つの実施形態によれば、音声処理用コンピュータプログラムが提供される。この音声処理用コンピュータプログラムは、第1の音声入力部により生成された第1の音声信号、及び、第1の音声入力部と異なる位置に配置された第2の音声入力部により生成された第2の音声信号を、それぞれ、所定の時間長を持つフレームごとに周波数領域の第1の周波数スペクトル及び第2の周波数スペクトルに変換し、フレームごとに、第1の周波数スペクトル及び第2の周波数スペクトルに基づいて、受音することが優先される第1の方向及び第1の方向と異なる第2の方向のうちの第2の方向に位置する音源のみが音声を発したる確からしさを算出し、フレームごとに、第1の周波数スペクトル及び第2の周波数スペクトルに基づいて算出される第1の方向から到来する音声を含む第1の指向音声信号を出力するとともに、確からしさに応じて、第1の周波数スペクトル及び第2の周波数スペクトルに基づいて算出される第2の方向から到来する音声を含む第2の指向音声信号を出力するか否かを制御する、ことをコンピュータに実行させるための命令を含む。

40

【発明の効果】

【0007】

一つの側面では、優先する方向に位置する音源からの音声だけでなく、他の方向に位置する他の音源からの音声も抑圧せずに出力できる。

50

【図面の簡単な説明】

【0008】

【図1】一つの実施形態による音声処理装置が実装された音声入力装置の概略構成図である。

【図2】音声処理装置の概略構成図である。

【図3】音声の到来方向と位相スペクトル差の関係の一例を示す図である。

【図4】第2の方向に位置する音源のみが音声を発した確からしさと第2の指向音声スペクトルに乗じるゲインとの関係の一例を示す図である。

【図5】受音についての指向特性を表す模式図である。

【図6】音声処理の動作フローチャートである。

10

【図7】変形例による、受音についての指向特性を表す模式図である。

【図8】第2の方向に位置する音源のみが音声を発した確からしさの程度が変化した時点からの経過時間と第1及び第2のゲインの関係の一例を示す図である。

【図9】変形例による指向特性制御部の指向特性制御の動作フローチャートである。

【図10】実施形態またはその変形例による音声処理装置の各部の機能を実現するコンピュータプログラムが動作することにより、音声処理装置として動作するコンピュータの構成図である。

【発明を実施するための形態】

【0009】

以下、図を参照しつつ、音声処理装置について説明する。この音声処理装置は、複数の音声入力部により得られた音声信号において、フレームごとに、優先される音源が位置する第1の方向と、他の音源が位置することが想定される第2の方向のうち、第2の方向に位置する音源のみが音声を発した確からしさを算出する。そしてこの音声処理装置は、その確からしさが高いフレームについて、第1の方向から到来する音声を含む第1の指向音声信号だけでなく、第2の方向から到来する音声を含む第2の指向音声信号も出力する。すなわち、この音声処理装置は、その確からしさが高いときに、受音する方向を一時的に第2の方向を含むように拡張する。

20

【0010】

図1は、一つの実施形態による音声処理装置が実装された音声入力装置の概略構成図である。音声入力装置1は、二つのマイクロホン11-1、11-2と、二つのアナログ/デジタル変換器12-1、12-2と、音声処理装置13と、通信インターフェース部14とを有する。音声入力装置1は、例えば、車両(図示せず)に搭載され、ドライバあるいは他の同乗者が発した音声を集音し、その音声を含む音声信号をナビゲーションシステム(図示せず)あるいはハンズフリーホン(図示せず)等へ出力する。そして音声処理装置13は、ドライバが位置する方向以外からの音声を抑圧するような受音の指向特性を設定する。さらに、音声処理装置13は、ドライバが位置する方向(第1の方向)と同乗者が位置する方向(第2の方向)のうち、同乗者のみが音声を発した確からしさが高い場合には、第2の方向から到来する音声も抑圧しないように指向特性を変化させる。

30

【0011】

マイクロホン11-1、11-2は、それぞれ、音声入力部の一例である。マイクロホン11-1及びマイクロホン11-2は、例えば、集音対象とする音源であるドライバと、他の音源である、助手席にいる同乗者(以下、単に同乗者と呼ぶ)との間において、例えば、インストゥルメントパネル、あるいは、車室内の天井付近に配置される。本実施形態では、マイクロホン11-1の方がマイクロホン11-2よりも同乗者に近く、かつ、マイクロホン11-2の方がマイクロホン11-1よりもドライバの近くに位置するように、マイクロホン11-1及びマイクロホン11-2は配置される。そしてマイクロホン11-1が周囲の音声を集音することにより生成したアナログの入力音声信号はアナログ/デジタル変換器12-1に入力される。同様に、マイクロホン11-2が周囲の音声を集音することにより生成したアナログの入力音声信号はアナログ/デジタル変換器12-2に入力される。

40

50

【0012】

アナログ/デジタル変換器12-1は、マイクロホン11-1から受け取ったアナログの入力音声信号を所定のサンプリング周波数でサンプリングすることによりデジタル化された入力音声信号を生成する。同様に、アナログ/デジタル変換器12-2は、マイクロホン11-2から受け取ったアナログの入力音声信号を所定のサンプリング周波数でサンプリングすることによりデジタル化された入力音声信号を生成する。

【0013】

なお、以下では、説明の便宜上、マイクロホン11-1が集音することで生成され、アナログ/デジタル変換器12-1によりデジタル化された入力音声信号を第1の入力音声信号と呼ぶ。また、マイクロホン11-2が集音することで生成され、アナログ/デジタル変換器12-2によりデジタル化された入力音声信号を第2の入力音声信号と呼ぶ。

10

アナログ/デジタル変換器12-1は、第1の入力音声信号を音声処理装置13へ出力する。同様に、アナログ/デジタル変換器12-2は、第2の入力音声信号を音声処理装置13へ出力する。

【0014】

音声処理装置13は、例えば、一つまたは複数のプロセッサと、メモリとを有する。そして音声処理装置13は、受信した第1の入力音声信号と第2の入力音声信号とから、制御される指向特性に応じて受音する方向以外の方向から到来した雑音を抑圧した指向音声信号を生成する。そして音声処理装置13は、通信インターフェース部14を介して、その指向音声信号をナビゲーションシステム(図示せず)あるいはハンズフリーホン(図示せず)といった他の機器へ出力する。

20

【0015】

通信インターフェース部14は、所定の通信規格に従って音声入力装置1を他の機器と接続するための通信インターフェース回路などを含む。例えば、通信インターフェース回路は、例えば、Bluetooth(登録商標)といった、音声信号の通信に利用可能な近距離無線通信規格に従って動作する回路、あるいは、universal serial bus(USB)といったシリアルバス規格に従って動作する回路とすることができる。そして通信インターフェース部14は、音声処理装置13から受け取った出力音声信号を他の機器へ出力する。

【0016】

図2は、一つの実施形態による音声処理装置13の概略構成図である。音声処理装置13は、時間周波数変換部21と、指向音声生成部22と、特徴抽出部23と、音源方向判定部24と、指向特性制御部25と、周波数時間変換部26とを有する。音声処理装置13が有するこれらの各部は、例えば、音声処理装置13が有するプロセッサ上で実行されるコンピュータプログラムによって実現される機能モジュールとして実装される。あるいは、音声処理装置13が有するこれらの各部は、音声処理装置13が有するプロセッサとは別個に、それらの各部の機能を実現する一つまたは複数の集積回路として音声処理装置13に実装されてもよい。

30

【0017】

時間周波数変換部21は、第1の入力音声信号及び第2の入力音声信号のそれぞれについて、フレーム単位で時間領域から周波数領域へ変換することにより、複数の周波数のそれぞれについての振幅成分と位相成分とを含む周波数スペクトルを算出する。なお、時間周波数変換部21は、第1の入力音声信号と第2の入力音声信号のそれぞれに対して同じ処理を行えばよいので、以下では、第1の入力音声信号についての処理について説明する。

40

【0018】

本実施形態では、時間周波数変換部21は、第1の入力音声信号を、所定のフレーム長(例えば、数10msec)を持つフレームごとに分割する。その際、時間周波数変換部21は、例えば、連続する二つのフレームがフレーム長の1/2だけずれるように各フレームを設定する。

【0019】

50

時間周波数変換部 2 1 は、各フレームに対して窓処理を実行する。すなわち、時間周波数変換部 2 1 は、各フレームに所定の窓関数を乗じる。例えば、時間周波数変換部 2 1 は、窓関数としてハニング窓を用いることができる。

【 0 0 2 0 】

時間周波数変換部 2 1 は、窓処理が施されたフレームを受け取る度に、そのフレームを時間領域から周波数領域へ変換することにより、複数の周波数のそれぞれについての振幅成分と位相成分とを含む周波数スペクトルを算出する。時間周波数変換部 2 1 は、例えば、フレームに対して、高速フーリエ変換(Fast Fourier Transform, FFT)といった時間周波数変換を実行することにより周波数スペクトルを算出すればよい。なお、以下では、便宜上、第 1 の入力音声信号について得られた周波数スペクトルを第 1 の周波数スペクトルと呼び、第 2 の入力音声信号について得られた周波数スペクトルを第 2 の周波数スペクトルと呼ぶ。

10

【 0 0 2 1 】

時間周波数変換部 2 1 は、フレームごとに、第 1 の周波数スペクトル及び第 2 の周波数スペクトルを指向音声生成部 2 2 へ出力する。

【 0 0 2 2 】

指向音声生成部 2 2 は、フレームごとに、マイクロホン 1 1 - 1 及び 1 1 - 2 から見て、受音することが優先される第 1 の方向(本実施形態では、ドライバが位置する方向)から到来する音声の周波数スペクトルを表す第 1 の指向音声スペクトルを生成する。また指向音声生成部 2 2 は、フレームごとに、マイクロホン 1 1 - 1 及び 1 1 - 2 から見て、他の音源が位置すると想定される第 2 の方向(本実施形態では、同乗者が位置する方向)から到来する音声の周波数スペクトルを表す第 2 の指向音声スペクトルを生成する。

20

【 0 0 2 3 】

先ず、指向音声生成部 2 2 は、例えば、フレームごとに、周波数ごとの第 1 の周波数スペクトルと第 2 の周波数スペクトル間の位相差を求める。この位相差は、そのフレームにおいて音声到来した方向に応じて変化するので、この位相差は、音声到来した方向を特定するために利用できる。例えば、位相差算出部 1 2 は、次式に従って周波数ごとの位相差を表す位相スペクトル差 $\Delta\theta(f)$ を求める。

【 数 1 】

$$\Delta\theta(f) = \tan^{-1}\left(\frac{IN1(f)}{IN2(f)}\right) \quad 0 < f < Fs/2 \quad (1)$$

30

ただし、 $IN1(f)$ は、第 1 の周波数スペクトルを表し、 $IN2(f)$ は、第 2 の周波数スペクトルを表す。そして f は周波数を表す。また Fs は、アナログ/デジタル変換器 1 2 - 1 及び 1 2 - 2 におけるサンプリング周波数を表す。

【 0 0 2 4 】

図 3 は、音声の到来方向と位相スペクトル差 $\Delta\theta(f)$ の関係の一例を示す図である。図 3 において、横軸は周波数を表し、縦軸は位相スペクトル差を表す。そして位相スペクトル差の範囲 3 0 1 は、第 1 の方向(本実施形態では、ドライバが位置する方向)から到来する音声第 1 の入力音声信号及び第 2 の入力音声信号に含まれる場合の周波数ごとの位相差の取り得る範囲を表す。一方、位相スペクトル差の範囲 3 0 2 は、第 2 の方向(本実施形態では、同乗者が位置する方向)から到来する音声第 1 の入力音声信号及び第 2 の入力音声信号に含まれる場合の周波数ごとの位相差の取り得る範囲を表す。

40

【 0 0 2 5 】

ドライバに対して、マイクロホン 1 1 - 2 の方がマイクロホン 1 1 - 1 よりも近い。そのため、ドライバが発した音声マイクロホン 1 1 - 1 に到達するタイミングがマイクロホン 1 1 - 2 に到達するタイミングよりも遅くなる。その結果として、第 1 の周波数スペ

50

クトルに表されるドライバが発した音声の位相は、第2の周波数スペクトルに表されるドライバが発した音声の位相よりも遅れる。そのため、位相スペクトル差の範囲301は、負側に位置する。そしてその遅れによる位相差の範囲は、周波数が高いほど広くなる。逆に、同乗者に対して、マイクロホン11-1の方がマイクロホン11-2よりも近い。そのため、同乗者が発した音声マイクロホン11-2に到達するタイミングがマイクロホン11-1に到達するタイミングよりも遅くなる。その結果として、第1の周波数スペクトルに表される同乗者が発した音声の位相は、第2の周波数スペクトルに表される同乗者が発した音声の位相よりも進む。そのため、位相スペクトル差の範囲302は、正側に位置する。そして位相差の範囲は、周波数が高いほど広くなる。

【0026】

そこで、指向音声生成部22は、各フレームについて、位相スペクトル差 (f)を参照して、周波数ごとに位相差が位相スペクトル差の範囲301に含まれるか、位相スペクトル差の範囲302に含まれるかを判定する。そして指向音声生成部22は、各フレームについて、第1及び第2の周波数スペクトルのうち、位相差が位相スペクトル差の範囲301に含まれる周波数の成分は、第1の方向から到来した音声に含まれる成分であると判定する。そして指向音声生成部22は、各フレームについて、第1の周波数スペクトルから、位相差が位相スペクトル差の範囲301に含まれる周波数の成分を抽出して第1の指向音声スペクトルとする。すなわち、指向音声生成部22は、位相差が位相スペクトル差の範囲301に含まれる周波数の成分に対して1となるゲインを乗じる。一方、指向音声生成部22は、位相差が位相スペクトル差の範囲301から外れる周波数の成分に対して0となるゲインを乗じる。これにより、指向音声生成部22は、第1の指向音声スペクトルを生成する。なお、指向音声生成部22は、位相スペクトル差の範囲301から外れる周波数の成分に対して、位相スペクトル差の範囲301から遠くなるほど小さくなるゲインを乗じてから、第1の指向音声スペクトルに含めてもよい。また、指向音声生成部22は、各フレームについて、第2の周波数スペクトルから、位相差が位相スペクトル差の範囲301に含まれる周波数の成分を抽出して第1の指向音声スペクトルとしてもよい。

【0027】

同様に、指向音声生成部22は、各フレームについて、第1及び第2の周波数スペクトルのうち、位相差が位相スペクトル差の範囲302に含まれる周波数の成分は、第2の方向から到来した音声に含まれる成分であると判定する。そして指向音声生成部22は、各フレームについて、第1の周波数スペクトルから、位相差が位相スペクトル差の範囲302に含まれる周波数の成分を抽出して第2の指向音声スペクトルとする。なお、指向音声生成部22は、位相スペクトル差の範囲302から外れる周波数の成分に対して、位相スペクトル差の範囲302から遠くなるほど小さくなるゲインを乗じてから、第2の指向音声スペクトルに含めてもよい。また、指向音声生成部22は、各フレームについて、第2の周波数スペクトルから、位相差が位相スペクトル差の範囲302に含まれる周波数の成分を抽出して第2の指向音声スペクトルとしてもよい。

【0028】

指向音声生成部22は、フレームごとに、第1の指向音声スペクトル及び第2の指向音声スペクトルのそれぞれを特徴抽出部23及び指向特性制御部25へ出力する。

【0029】

特徴抽出部23は、フレームごとに、第1及び第2の指向音声スペクトルに基づいて、そのフレームについて音源からの音声らしさを表す特徴量を算出する。

【0030】

第1の方向に位置する音源（この例では、ドライバ）が発した音声が含まれるフレームについて、第1の方向からの音声が大きくなるので、第1の指向音声スペクトルのパワーはある程度大きくなると想定される。同様に、第2の方向に位置する音源（この例では、同乗者）が発した音声が含まれるフレームについて、第2の方向からの音声が大きくなるので、第2の指向音声スペクトルのパワーはある程度大きくなると想定される。また、ドライバの音声のパワー及び同乗者の音声のパワーは経時変化すると想定される。そこで、

10

20

30

40

50

本実施形態では、特徴抽出部 2 3 は、フレームごとに、第 1 及び第 2 の指向音声スペクトルのそれぞれについて、特徴量として、パワーと、パワーについての非定常性度合い（以下、単に非定常性度と呼ぶ）とを算出する。

【 0 0 3 1 】

例えば、特徴抽出部 2 3 は、次式に従って、フレームごとに、第 1 の指向音声スペクトルのパワー PX 及び第 2 の指向音声スペクトルのパワー PY を算出する。

【数 2】

$$PX = \sum_f X(f)^2 \quad (2) \quad 10$$

$$PY = \sum_f Y(f)^2$$

ここで、 $X(f)$ は、着目するフレームについての第 1 の指向音声スペクトルであり、 $Y(f)$ は、着目するフレームについての第 2 の指向音声スペクトルである。

【 0 0 3 2 】

また、特徴抽出部 2 3 は、次式に従って、フレームごとに、第 1 の指向音声スペクトルの非定常性度 RX 及び第 2 の指向音声スペクトルの非定常性度 RY を算出する。

【数 3】

$$RX = \left| 10 \times \log_{10} (PX/PX') \right| \quad (3)$$

$$RY = \left| 10 \times \log_{10} (PY/PY') \right|$$

ここで、 PX' は、着目するフレームの一つ前のフレームについての第 1 の指向音声スペクトルのパワーを表し、 PY' は、着目するフレームの一つ前のフレームについての第 2 の指向音声スペクトルのパワーを表す。

特徴抽出部 2 3 は、フレームごとに、算出した特徴量を音源方向判定部 2 4 へわたす。

【 0 0 3 3 】

音源方向判定部 2 4 は、フレームごとに、第 1 の指向音声スペクトルの特徴量と第 2 の指向音声スペクトルの特徴量とに基づいて、そのフレームにおいて、第 1 の方向と第 2 の方向のうち、第 2 の方向に位置する音源のみが音声を発した確からしさを判定する。以下では、第 1 の方向と第 2 の方向のうち、第 2 の方向に位置する音源のみが音声を発した確からしさを、単に第 2 の方向に位置する音源のみが音声を発した確からしさと呼ぶ。

【 0 0 3 4 】

上記のように、第 1 の方向に位置する音源が発した音声が含まれるフレームについて、第 1 の指向音声スペクトルのパワー及び非定常性度はある程度大きくなると想定される。一方、第 2 の方向に位置する音源が発した音声が含まれるフレームについて、第 2 の指向音声スペクトルのパワー及び非定常性度はある程度大きくなると想定される。したがって、音源方向判定部 2 4 は、フレームごとに、第 2 の方向に位置する音源のみが音声を発した確からしさ P を、次式に従って算出する。

10

20

30

40

【数4】

$$P = \frac{PY}{PX} + \frac{RY}{RX} \quad (4)$$

【0035】

したがって、確からしさPの値が大きいほど、第1の方向及び第2の方向のうち、第2の方向に位置する音源のみが音声を発している可能性が高い。音源方向判定部24は、フレームごとに、第2の方向に位置する音源のみが音声を発した確からしさPを、指向特性制御部25へ通知する。

10

【0036】

指向特性制御部25は、周波数時間変換部26とともに、指向音声出力部の一例を形成する。そして指向特性制御部25は、フレームごとに、第2の方向に位置する音源のみが音声を発した確からしさに応じて、受信する指向特性を制御する。本実施形態では、指向特性制御部25は、第1の指向音声スペクトルを常に出力し、第2の指向音声スペクトルには抑圧の程度を表すゲインを乗じて出力する。そして指向特性制御部25は、そのゲインを、確からしさPに応じて制御する。

20

【0037】

本実施形態では、指向特性制御部25は、フレームごとに、算出した確からしさPを少なくとも一つの尤度判定閾値と比較する。例えば、指向特性制御部25は、着目するフレームについて、確からしさPが第1の尤度判定閾値Th1よりも高い場合、そのフレームにおいて第2の方向に位置する音源のみが音声を発した確からしさが高いと判定する。一方、指向特性制御部25は、着目するフレームについて、確からしさPが第2の尤度判定閾値Th2(ただし、Th2<Th1)よりも低い場合、そのフレームにおいて第2の方向に位置する音源のみが音声を発した確からしさは低いと判定する。また、着目するフレームについて、確からしさPが第2の尤度判定閾値Th2以上、かつ、第1の尤度判定閾値Th1以下であれば、音源方向判定部24は、そのフレームにおいて第2の方向に位置する音源のみが音声を発した確からしさは中程度であると判定する。

30

【0038】

着目するフレームについて、第2の方向に位置する音源のみが音声を発した確からしさが低い場合、指向特性制御部25は、第1の指向音声スペクトル及び第2の指向音声スペクトルのうち、第1の指向音声スペクトルのみを出力する。すなわち、指向特性制御部25は、第2の指向音声スペクトルに乗じるゲインを0に設定することで、受信する指向特性を第1の方向に制限する。一方、着目するフレームについて、第2の方向に位置する音源のみが音声を発した確からしさが高い場合、指向特性制御部25は、第1の指向音声スペクトル及び第2の指向音声スペクトルの両方を出力する。すなわち、指向特性制御部25は、第2の指向音声スペクトルに乗じるゲインを1に設定することで、受信する指向特性を、第1の方向だけでなく、第2の方向にも拡張する。

40

【0039】

また、着目するフレームについて、第2の方向に位置する音源のみが音声を発した確からしさの程度が中程度である場合、指向特性制御部25は、第2の指向音声スペクトルに乗じるゲインを、確からしさPの値が高くなるほど1に近くなるように決定する。

【0040】

図4は、第2の方向に位置する音源のみが音声を発した確からしさPと第2の指向音声スペクトルに乗じるゲインGとの関係の一例を示す図である。図4において、横軸は確からしさPを表し、縦軸は、ゲインGを表す。そしてグラフ400は、確からしさPとゲインGの関係を表す。

50

【 0 0 4 1 】

グラフ 4 0 0 に示されるように、確からしさPが第 2 の尤度判定閾値Th2以下である場合、ゲインGは0に設定される。また、確からしさPが第 1 の尤度判定閾値Th1以上である場合、ゲインGは1に設定される。そして確からしさPが第 2 の尤度判定閾値Th2よりも大きく、かつ、第 1 の尤度判定閾値Th1未満である場合、確からしさPが高くなるにつれてゲインGも単調かつ線形に高くなる。

【 0 0 4 2 】

なお、変形例によれば、一つの尤度判定閾値Thが用いられてもよい。この場合には、着目するフレームについて、確からしさPが尤度判定閾値Thよりも高い場合、指向特性制御部 2 5 は、そのフレームにおいて第 2 の方向に位置する音源のみが音声を発した確からしさが高いと判定する。一方、確からしさPが尤度判定閾値Th以下である場合、指向特性制御部 2 5 は、そのフレームにおいて第 2 の方向に位置する音源のみが音声を発した確からしさが低いと判定する。

10

【 0 0 4 3 】

なお、尤度判定閾値Th1、Th2、Thは、例えば、実験などにより予め設定され、音声処理装置 1 3 が有するメモリに予め保存されればよい。

【 0 0 4 4 】

図 5 は、受音についての指向特性を表す模式図である。第 2 の方向に位置する音源のみが音声を発した確からしさの程度が低い場合、受音する感度が高い範囲 5 0 1 は、マイクロホン 1 1 - 1 とマイクロホン 1 1 - 2 の並び方向について、ドライバ 5 1 1 が位置するマイクロホン 1 1 - 2 側に設定される。一方、第 2 の方向に位置する音源のみが音声を発した確からしさの程度が高い場合、受音する感度が高い範囲 5 0 2 は、マイクロホン 1 1 - 1 とマイクロホン 1 1 - 2 の並び方向について、マイクロホン 1 1 - 2 側とともに、マイクロホン 1 1 - 1 側にも設定される。これにより、ドライバ 5 1 1 が位置する方向だけでなく、同乗者 5 1 2 が位置する方向も受音する感度が高い範囲に含まれる。

20

【 0 0 4 5 】

周波数時間変換部 2 6 は、フレームごとに、指向特性制御部 2 5 から出力された第 1 の指向音声スペクトルを、周波数時間変換して時間領域の信号に変換することにより、フレームごとの第 1 の指向音声信号を得る。また、周波数時間変換部 2 6 は、フレームごとに、指向特性制御部 2 5 から出力された第 2 の指向音声スペクトルを、周波数時間変換して時間領域の信号に変換することにより、フレームごとの第 2 の指向音声信号を得る。なお、この周波数時間変換は、時間周波数変換部 2 1 により行われる時間周波数変換の逆変換である。

30

【 0 0 4 6 】

周波数時間変換部 2 6 は、時間順（すなわち、再生順）に連続するフレームごとの第 1 の指向音声信号を、フレーム長の1/2ずつずらして加算することにより、第 1 の指向音声信号を算出する。同様に、周波数時間変換部 2 6 は、時間順に連続するフレームごとの第 2 の指向音声信号を、フレーム長の1/2ずつずらして加算することにより、第 2 の指向音声信号を算出する。そして周波数時間変換部 2 6 は、第 1 の指向音声信号及び第 2 の指向音声信号を、通信インターフェース部 1 4 を介して他の機器へ出力する。

40

【 0 0 4 7 】

図 6 は、音声処理装置 1 3 により実行される音声処理の動作フローチャートである。音声処理装置 1 3 は、フレームごとに、下記のフローチャートに従って音声処理を実行する。

【 0 0 4 8 】

時間周波数変換部 2 1 は、フレーム単位に分割された第 1 の入力音声信号及び第 2 の入力音声信号にハニング窓関数を乗じる（ステップ S 1 0 1）。そして、時間周波数変換部 2 1 は、第 1 の入力音声信号及び第 2 の入力音声信号を時間周波数変換して第 1 の周波数スペクトル及び第 2 の周波数スペクトルを算出する（ステップ S 1 0 2）。

【 0 0 4 9 】

50

指向音声生成部 2 2 は、第 1 及び第 2 の周波数スペクトルに基づいて、第 1 の指向音声スペクトル及び第 2 の指向音声スペクトルを生成する（ステップ S 1 0 3）。特徴抽出部 2 3 は、音源からの音声らしさを表す特徴量として、第 1 の指向音声スペクトルのパワー及び非定常性度と、第 2 の指向音声スペクトルのパワー及び非定常性度を算出する（ステップ S 1 0 4）。

【 0 0 5 0 】

音源方向判定部 2 4 は、第 1 の指向音声スペクトル及び第 2 の指向音声スペクトルのそれぞれのパワー及び非定常性度に基づいて、第 1 及び第 2 の方向のうち、第 2 の方向に位置する音源のみから音声到来する確からしさ P を算出する（ステップ S 1 0 5）。

【 0 0 5 1 】

指向特性制御部 2 5 は、確からしさ P が第 1 の尤度判定閾値 Th1 よりも大きいかが判定する（ステップ S 1 0 6）。確からしさ P が第 1 の尤度判定閾値 Th1 より大きい場合（ステップ S 1 0 6 - Y e s）、指向特性制御部 2 5 は、第 1 及び第 2 の指向音声スペクトルの両方を出力する（ステップ S 1 0 7）。一方、確からしさ P が第 1 の尤度判定閾値 Th1 以下である場合（ステップ S 1 0 6 - N o）、指向特性制御部 2 5 は、確からしさ P が第 2 の尤度判定閾値 Th2 よりも小さいかが判定する（ステップ S 1 0 8）。確からしさ P が第 2 の尤度判定閾値 Th2 よりも小さい場合（ステップ S 1 0 8 - Y e s）、指向特性制御部 2 5 は、第 1 及び第 2 の指向音声スペクトルのうちの第 1 の指向音声スペクトルのみを出力する（ステップ S 1 0 9）。すなわち、指向特性制御部 2 5 は、第 1 の指向音声スペクトルとともに、振幅が全周波数帯域にわたって 0 となる第 2 の指向音声スペクトルを出力する。一方、確からしさ P が第 2 の尤度判定閾値 Th2 以上である場合（ステップ S 1 0 8 - N o）、指向特性制御部 2 5 は、第 1 の指向音声スペクトルとともに、確からしさ P に応じて抑圧した第 2 の指向音声スペクトルを出力する（ステップ S 1 1 0）。

【 0 0 5 2 】

周波数時間変換部 2 6 は、指向特性制御部 2 5 から出力された第 1 の指向音声スペクトルを周波数時間変換して第 1 の指向音声信号を算出する。また周波数時間変換部 2 6 は、第 2 の指向音声スペクトルが出力された場合には、第 2 の指向音声スペクトルについても周波数時間変換して第 2 の指向音声信号を算出する（ステップ S 1 1 1）。そして周波数時間変換部 2 6 は、前フレームまでの第 1 の指向音声信号に対して半フレーム長ずらして現フレームの第 1 の指向音声信号を合成する。同様に、周波数時間変換部 2 6 は、前フレームまでの第 2 の指向音声信号に対して半フレーム長ずらして現フレームの第 2 の指向音声信号を合成する（ステップ S 1 1 2）。そして音声処理装置 1 3 は、音声処理を終了する。

【 0 0 5 3 】

以上に説明してきたように、この音声処理装置は、受音することが優先される音源が位置する第 1 の方向と、他の音源が位置することが想定される第 2 の方向のうち第 2 の方向に位置する音源のみが音声を発した確からしさをフレームごとに算出する。そしてこの音声処理装置は、その確からしさが高いと、第 1 の方向から到来する音声を含む第 1 の指向音声信号だけでなく、第 2 の方向から到来する音声を含む第 2 の指向音声信号も出力する。すなわち、この音声処理装置は、その確からしさが高いと、受音の指向特性を、第 1 の方向だけでなく、第 2 の方向も含むように制御する。これにより、この音声処理装置は、例えば、複数の話者のうちの特定の話者が発した音声を優先的に受音しつつ、他の話者が音声を発したときには、他の話者が発した音声も受音することを可能とする。

【 0 0 5 4 】

なお、変形例によれば、特徴抽出部 2 3 は、フレームごとに、音源からの音声らしさを表す特徴量として、第 1 の指向音声スペクトルのパワーと、第 2 の指向音声スペクトルのパワーを算出し、非定常性度については算出しなくてもよい。この場合には、特徴抽出部 2 3 は、確からしさ P を、次式に従って算出すればよい。

10

20

30

40

【数5】

$$P = \frac{PY}{PX} \quad (5)$$

【0055】

また他の変形例によれば、指向音声生成部22は、第1の周波数スペクトルと第2の周波数スペクトル間の同期減算により、フレームごとに第1の指向音声スペクトル及び第2の指向音声スペクトルを算出してもよい。この場合、指向音声生成部22は、次式に従って第1の指向音声スペクトル $X(f)$ 及び第2の指向音声スペクトル $Y(f)$ を算出する。

10

【数6】

$$\begin{aligned} X(f) &= IN1(f) - e^{-j2\pi fn/N} IN2(f) \\ Y(f) &= IN2(f) - e^{-j2\pi fn/N} IN1(f) \end{aligned} \quad (6)$$

20

ここで、 N は、1フレームに含まれるサンプリング点の総数、すなわち、フレーム長を表す。また n は、マイクロホン11-1とマイクロホン11-2間の、音源から音声到達するサンプリング時間差を表す。なお、 n が $0 < n < 1$ 、すなわち、サンプリング間隔以下となるように、マイクロホン11-1とマイクロホン11-2間の間隔 d は、(音速/ F_s)以下となるように設定される。

【0056】

図7は、この変形例による、受音についての指向特性を表す模式図である。第2の方向に位置する音源のみが音声を発した確からしさの程度が低い場合、受音する感度が高い範囲701は、マイクロホン11-1とマイクロホン11-2の並び方向について、ドライバ711が位置するマイクロホン11-2側に設定される。一方、第2の方向に位置する音源のみが音声を発した確からしさの程度が高い場合、受音する感度が高い範囲702は、マイクロホン11-2側とともに、同乗者712が位置するマイクロホン11-1側にも設定される。またこの例では、第1の指向音声信号について受音する感度が高い範囲と、第2の指向音声信号について受音する感度が高い範囲の一部が重なる。

30

【0057】

さらに他の変形例によれば、指向特性制御部25は、フレームごとに、第1の指向音声スペクトルに抑圧の程度を表す第1のゲインを乗じて得られるスペクトルを出力してもよい。同様に、指向特性制御部25は、フレームごとに、第2の指向音声スペクトルに抑圧の程度を表す第2のゲインを乗じて得られるスペクトルを出力してもよい。そして指向特性制御部25は、第2の方向に位置する音源のみが音声を発した確からしさの程度が変化した時点からの経過時間に応じて、第1のゲイン及び第2のゲインを調節してもよい。

40

【0058】

図8は、第2の方向に位置する音源のみが音声を発した確からしさの程度が変化した時点からの経過時間と第1及び第2のゲインの関係の一例を示す図である。図8において、横軸は時間を表し、縦軸はゲインを表す。そしてグラフ801は、第2の方向に位置する音源のみが音声を発した確からしさの程度が変化した時点からの経過時間と第1のゲインの関係を表す。またグラフ802は、第2の方向に位置する音源のみが音声を発した確からしさの程度が変化した時点からの経過時間と第2のゲインの関係を表す。

【0059】

この例では、時刻 t_1 までは、第2の方向に位置する音源のみが音声を発した確からしさ

50

Pが第1の尤度判定閾値 $Th1$ 以下であり、時刻 $t1$ において確からしさPが第1の尤度判定閾値 $Th1$ より大きくなったとする。すなわち、時刻 $t1$ において、第2の方向に位置する音源のみが音声を発した確からしさの程度が高いに変化したとする。また、時刻 $t1$ 以降、時刻 $t3$ までは、第2の方向に位置する音源のみが音声を発した確からしさPは第2の尤度判定閾値 $Th2$ 以上であり、時刻 $t3$ において確からしさPが第2の尤度判定閾値 $Th2$ より小さくなったとする。すなわち、時刻 $t3$ において、第2の方向に位置する音源のみが音声を発した確からしさの程度が低いに変化したとする。

【0060】

この場合、時刻 $t1$ までは、第1のゲイン $G1$ は1に設定され、一方、第2のゲイン $G2$ は0に設定される。すなわち、第2の方向に位置する音源のみが音声を発した確からしさの程度が高いに変化するまでは、指向特性制御部25は、第1の指向音声スペクトルをそのまま出力し、かつ、第2の指向音声スペクトルを出力しない。

10

【0061】

一方、時刻 $t1$ になり、第2の方向に位置する音源のみが音声を発した確からしさの程度が高いに変化すると、その後の時刻 $t2$ までの一定期間（例えば、数10msec）、指向特性制御部25は、第1のゲイン $G1$ を線形に単調減少させる。そして時刻 $t2$ 以降、指向特性制御部25は、第1のゲイン $G1$ を、 $0 < G1 < 1$ となる所定の値（この例では、0.7）に設定する。一方、指向特性制御部25は、時刻 $t1$ 以降、第2のゲイン $G2$ を1に設定する。すなわち、指向特性制御部25は、第1の指向音声スペクトルを減衰させて出力し、かつ、第2の指向音声スペクトルをそのまま出力する。これにより、第2の方向に位置する音源から音声が到来している間は、第2の指向音声信号に含まれる、第2の方向からの音声についての、第1の方向から受信した雑音に対する信号対雑音比が向上する。

20

【0062】

また、時刻 $t3$ になり、第2の方向に位置する音源のみが音声を発した確からしさの程度が低いに変化すると、指向特性制御部25は、その後の時刻 $t4$ までの一定期間（例えば、100msec～200msec）は第1のゲイン $G1$ を所定値に維持する。そして指向特性制御部25は、時刻 $t4$ 以降、第1のゲイン $G1$ を1に戻す。また、指向特性制御部25は、時刻 $t4$ まで、第2のゲイン $G2$ を1に維持し、時刻 $t4$ 以降、第2のゲイン $G2$ を線形に単調減少させる。そして指向特性制御部25は、時刻 $t4$ よりも後の時刻 $t5$ 以降、第2のゲイン $G2$ を0にする。これにより、第2の方向に位置する音源のみが音声を発した確からしさの程度が低いに変化しても、その後の一定期間の間、第2の指向音声スペクトルは出力される。そのため、例えば、第2の指向音声信号に含まれる、第2の方向からの音声の後端部分、例えば、第2の方向に位置する同乗者が発した会話音声の語尾部分が途切れることが防止される。したがって、例えば、第2の指向音声信号を受信した他の機器が、第2の指向音声信号から同乗者の音声を認識する場合、語尾部分が途切れることによる認識精度の低下が防止される。なお、時刻 $t3$ ～時刻 $t5$ までの期間は、時刻 $t3$ ～時刻 $t4$ までの期間以上であり、かつ、例えば、100msec～300msecに設定される。

30

【0063】

図9は、この変形例による指向特性制御部25の指向特性制御の動作フローチャートである。なお、この指向特性制御の処理は、図6に示される音声処理の動作フローチャートにおけるステップS106～S110までの処理の代わりに実行される。また図9では、現フレームにおける、第2の方向に位置する音源のみが音声を発した確からしさを $P(t)$ と表記し、直前のフレームにおける、第2の方向に位置する音源のみが音声を発した確からしさを $P(t-1)$ と表記する。

40

【0064】

図6に示されたステップS105において、現フレームの確からしさ $P(t)$ が算出されると、指向特性制御部25は、確からしさ $P(t)$ が第1の尤度判定閾値 $Th1$ よりも大きいかが判定する（ステップS201）。確からしさ $P(t)$ が第1の尤度判定閾値 $Th1$ よりも大きい場合（ステップS201 - Yes）、指向特性制御部25は、直前のフレームの確からしさ $P(t-1)$ が第1の尤度判定閾値 $Th1$ 以下か否かが判定する（ステップS202）。確から

50

しさを $P(t-1)$ が第1の尤度判定閾値 $Th1$ 以下であれば(ステップS202 - Yes)、現フレームにおいて、第2の方向に位置する音源のみが音声を発した確からしさが高いに変化している。そこで、指向特性制御部25は、第2の方向に位置する音源のみが音声を発した確からしさが高いに変化してからの経過時間を表すフレーム数 $cnt1$ を1に設定する。また、指向特性制御部25は、第2の方向に位置する音源のみが音声を発した確からしさが低いに変化してからの経過時間を表すフレーム数 $cnt2$ を0に設定する(ステップS203)。なお、初期状態では、第1のゲイン $G1$ が1、第2のゲイン $G2$ が0となるように、フレーム数 $cnt1$ は0に設定され、かつ、フレーム数 $cnt2$ は、時刻 $t3$ ~時刻 $t5$ の期間に相当するフレーム数よりも大きい値に設定される。

【0065】

一方、確からしさを $P(t-1)$ が第1の尤度判定閾値 $Th1$ よりも高ければ(ステップS202 - No)、直前のフレームの時点でも、第2の方向に位置する音源のみが音声を発した確からしさが高く、その確からしさが高い状態が現フレームまで継続している。そのため、指向特性制御部25は、フレーム数 $cnt1$ を1インクリメントする(ステップS204)。そしてステップS203またはS204の後、指向特性制御部25は、第1のゲイン $G1$ を、例えば、図8に示されるように、フレーム数 $cnt1$ に応じて設定し、第2のゲイン $G2$ を1に設定する(ステップS205)。

【0066】

また、ステップS201において、確からしさを $P(t)$ が第1の尤度判定閾値 $Th1$ 以下である場合(ステップS201 - No)、指向特性制御部25は、 $P(t)$ が第2の尤度判定閾値 $Th2$ よりも小さいか否か判定する(ステップS206)。 $P(t)$ が第2の尤度判定閾値 $Th2$ よりも小さい場合(ステップS206 - Yes)、指向特性制御部25は、直前のフレームの確からしさを $P(t-1)$ が第2の尤度判定閾値 $Th2$ 以上か否か判定する(ステップS207)。確からしさを $P(t-1)$ が第2の尤度判定閾値 $Th2$ 以上であれば(ステップS207 - Yes)、現フレームにおいて、第2の方向に位置する音源のみが音声を発した確からしさが低いに変化している。そこで、指向特性制御部25は、フレーム数 $cnt1$ を0に設定し、かつ、フレーム数 $cnt2$ を1に設定する(ステップS208)。

【0067】

一方、確からしさを $P(t-1)$ が第2の尤度判定閾値 $Th2$ よりも低ければ(ステップS207 - No)、直前のフレームの時点でも、第2の方向に位置する音源のみが音声を発した確からしさが低く、その確からしさが低い状態が現フレームまで継続している。そのため、指向特性制御部25は、フレーム数 $cnt2$ を1インクリメントする(ステップS209)。そしてステップS208またはS209の後、指向特性制御部25は、第1のゲイン $G1$ 及び第2のゲイン $G2$ を、例えば、図8に示されるように、フレーム数 $cnt2$ に応じて設定する(ステップS210)。

【0068】

また、ステップS206にて、 $P(t)$ が第2の尤度判定閾値 $Th2$ 以上である場合(ステップS206 - No)、現フレームでは、確からしさが中程度の状態であることが継続している。そこで、指向特性制御部25は、フレーム数 $cnt1$ が0よりも大きいかが判定する(ステップS211)。フレーム数 $cnt1$ が0よりも大きければ(ステップS211 - Yes)、確からしさが高い状態が継続しているとみなす。そこで指向特性制御部25は、フレーム数 $cnt1$ を1インクリメントする(ステップS204)。一方、フレーム数 $cnt1$ が0であれば(ステップS211 - No)、フレーム数 $cnt2$ が0よりも大きいはずなので、確からしさが低い状態が継続しているとみなす。そこで指向特性制御部25は、フレーム数 $cnt2$ を1インクリメントする(ステップS209)。

【0069】

ステップS205またはステップS210の後、指向特性制御部25は、第1のゲイン $G1$ を第1の指向音声スペクトルに乗じてからその第1の指向音声スペクトルを出力する。また、指向特性制御部25は、第2のゲイン $G2$ を第2の指向音声スペクトルに乗じてからその第2の指向音声スペクトルを出力する(ステップS212)。そして音声処理装置1

10

20

30

40

50

3は、図6のステップS111以降の処理を実行する。

【0070】

この変形例によれば、音声処理装置は、第2の方向に位置する音源のみが音声を発している場合のその音声についての信号対雑音比を向上できるとともに、第2の方向に位置する音源から発した音声の語尾が途切れることを防止できる。

なお、この変形例においても、二つの第1の尤度判定閾値 $Th1$ と第2の尤度判定閾値 $Th2$ の代わりに、一つの尤度判定閾値 Th が用いられてもよい。この場合には、指向特性制御部25は、図9に示された動作フローチャートにおいて、 $Th1=Th2=Th$ として、指向特性制御を行えばよい。

【0071】

上記の実施形態または変形例において、指向特性制御部25は、フレームごとに、ゲインが乗じられた後の第1の指向音声スペクトルと第2の指向音声スペクトルを合成して一つのスペクトルとしてから出力してもよい。そして周波数時間変換部26は、その一つのスペクトルを周波数時間変換してフレームごとに合成することで、一つの指向音声信号を算出し、その指向音声信号を出力してもよい。あるいは、周波数時間変換部26は、第1の指向音声信号と第2の指向音声信号を合成して一つの指向音声信号を算出し、その指向音声信号を出力してもよい。

【0072】

上記の実施形態または変形例による音声処理装置は、上記のような音声入力装置以外の装置、例えば、電話会議システムなどに実装されてもよい。

【0073】

上記の実施形態または変形例による音声処理装置が有する各機能をコンピュータに実現させるコンピュータプログラムは、磁気記録媒体あるいは光記録媒体といった、コンピュータによって読み取り可能な媒体に記録された形で提供されてもよい。

【0074】

図10は、上記の実施形態またはその変形例による音声処理装置の各部の機能を実現するコンピュータプログラムが動作することにより、音声処理装置として動作するコンピュータの構成図である。

コンピュータ100は、ユーザインターフェース部101と、オーディオインターフェース部102と、通信インターフェース部103と、記憶部104と、記憶媒体アクセス装置105と、プロセッサ106とを有する。プロセッサ106は、ユーザインターフェース部101、オーディオインターフェース部102、通信インターフェース部103、記憶部104及び記憶媒体アクセス装置105と、例えば、バスを介して接続される。

【0075】

ユーザインターフェース部101は、例えば、キーボードとマウスなどの入力装置と、液晶ディスプレイといった表示装置とを有する。または、ユーザインターフェース部101は、タッチパネルディスプレイといった、入力装置と表示装置とが一体化された装置を有してもよい。そしてユーザインターフェース部101は、例えば、ユーザの操作に応じて、音声処理を開始させる操作信号をプロセッサ106へ出力する。

【0076】

オーディオインターフェース部102は、コンピュータ100を、マイクロホン（図示せず）と接続するためのインターフェース回路を有する。そしてオーディオインターフェース部102は、2以上のマイクロホンのそれぞれから受け取った入力音声信号をプロセッサ106へ渡す。

【0077】

通信インターフェース部103は、イーサネット（登録商標）などの通信規格に従った通信ネットワークに接続するための通信インターフェース及びその制御回路を有する。そして通信インターフェース部103は、例えば、プロセッサ106から受け取った、第1の指向音声信号及び第2の指向音声信号のそれぞれを通信ネットワークを介して他の機器へ出力する。あるいは、通信インターフェース部103は、第1の指向音声信号及び第2

10

20

30

40

50

の指向音声信号に対して音声認識処理を適用することで得られた音声認識結果を、通信ネットワークを介して他の機器へ出力してもよい。あるいはまた、通信インターフェース部103は、音声認識結果に応じて実行されたアプリケーションにより生成された信号を、通信ネットワークを介して他の機器へ出力してもよい。

【0078】

記憶部104は、例えば、読み書き可能な半導体メモリと読み出し専用の半導体メモリとを有する。そして記憶部104は、プロセッサ106上で実行される、音声処理を実行するためのコンピュータプログラム、及び音声処理で利用される様々なデータまたは音声処理の途中で生成される各種の信号などを記憶する。

【0079】

記憶媒体アクセス装置105は、例えば、磁気ディスク、半導体メモリカード及び光記憶媒体といった記憶媒体107にアクセスする装置である。記憶媒体アクセス装置105は、例えば、記憶媒体107に記憶された、プロセッサ106上で実行される音声処理用のコンピュータプログラムを読み込み、プロセッサ106に渡す。

【0080】

プロセッサ106は、上記の実施形態または変形例による音声処理用コンピュータプログラムを実行することにより、各入力音声信号から第1の指向音声信号及び第2の指向音声信号を生成する。そしてプロセッサ106は、第1の指向音声信号及び第2の指向音声信号を通信インターフェース部103へ出力する。

【0081】

さらに、プロセッサ106は、第1の指向音声信号に対して音声認識処理を実行することで、第1の方向に位置する話者が発した音声を認識してもよい。同様に、プロセッサ106は、第2の指向音声信号に対して音声認識処理を実行することで、第2の方向に位置する他の話者が発した音声を認識してもよい。そしてプロセッサ106は、それぞれの音声認識結果に応じて所定のアプリケーションを実行してもよい。

【0082】

ここに挙げられた全ての例及び特定の用語は、読者が、本発明及び当該技術の促進に対する本発明者により寄与された概念を理解することを助ける、教示的な目的において意図されたものであり、本発明の優位性及び劣等性を示すことに関する、本明細書の如何なる例の構成、そのような特定の挙げられた例及び条件に限定しないように解釈されるべきものである。本発明の実施形態は詳細に説明されているが、本発明の精神及び範囲から外れることなく、様々な変更、置換及び修正をこれに加えることが可能であることを理解されたい。

【0083】

以上説明した実施形態及びその変形例に関し、更に以下の付記を開示する。

(付記1)

第1の音声入力部により生成された第1の音声信号、及び、前記第1の音声入力部と異なる位置に配置された第2の音声入力部により生成された第2の音声信号を、それぞれ、所定の時間長を持つフレームごとに周波数領域の第1の周波数スペクトル及び第2の周波数スペクトルに変換し、

フレームごとに、前記第1の周波数スペクトル及び前記第2の周波数スペクトルに基づいて、受音することが優先される第1の方向及び前記第1の方向と異なる第2の方向のうちの前記第2の方向に位置する音源のみが音声を発した確からしさを算出し、

フレームごとに、前記第1の周波数スペクトル及び前記第2の周波数スペクトルに基づいて算出される前記第1の方向から到来する音声を含む第1の指向音声信号を出力するとともに、前記確からしさに応じて、前記第1の周波数スペクトル及び前記第2の周波数スペクトルに基づいて算出される前記第2の方向から到来する音声を含む第2の指向音声信号を出力するか否かを制御する、

ことをコンピュータに実行させるための音声処理用コンピュータプログラム。

(付記2)

10

20

30

40

50

前記第 2 の指向音声信号の出力を制御することは、前記確からしさが第 1 の閾値よりも高くなるフレームについて前記第 2 の指向音声信号を出力する、付記 1 に記載の音声処理用コンピュータプログラム。

(付記 3)

前記第 2 の指向音声信号の出力を制御することは、第 1 のフレームにおける前記確からしさが前記第 1 の閾値よりも低い第 2 の閾値未満となり、かつ、前記第 1 のフレームの直前のフレームにおける前記確からしさが前記第 2 の閾値以上である場合、前記第 1 のフレームから第 1 の期間経過後のフレームから前記第 2 の指向音声信号の出力を停止する、付記 2 に記載の音声処理用コンピュータプログラム。

(付記 4)

前記第 2 の指向音声信号の出力を制御することは、第 2 のフレームにおける前記確からしさが前記第 1 の閾値よりも高く、かつ、前記第 2 のフレームの直前のフレームにおける前記確からしさが前記第 1 の閾値以下である場合、前記第 2 のフレームから第 2 の期間にわたって前記第 1 の指向音声信号を抑圧して出力する、付記 3 に記載の音声処理用コンピュータプログラム。

(付記 5)

前記第 2 の指向音声信号の出力を制御することは、前記第 2 のフレーム以降の第 3 のフレームにおける前記確からしさが前記第 2 の閾値未満となる場合、前記第 3 のフレームから第 3 の期間経過した時点を前記第 2 の期間の終端とする、付記 4 に記載の音声処理用コンピュータプログラム。

(付記 6)

フレームごとに、前記第 1 の周波数スペクトル及び前記第 2 の周波数スペクトルに基づいて、前記第 1 の指向音声信号のパワー及び前記第 2 の指向音声信号のパワーを算出することをさらにコンピュータに実行させ、

前記確からしさを算出することは、フレームごとに、前記第 1 の指向音声信号のパワーに対する前記第 2 の指向音声信号のパワーのパワー比に基づいて前記確からしさを算出する、付記 1 ~ 5 の何れかに記載の音声処理用コンピュータプログラム。

(付記 7)

フレームごとに、前記第 1 の周波数スペクトル及び前記第 2 の周波数スペクトルに基づいて、前記第 1 の指向音声の非定常性度合い及び前記第 2 の指向音声の非定常性度合いを算出することをさらにコンピュータに実行させ、

前記確からしさを算出することは、フレームごとに、前記第 1 の指向音声の非定常性度合いに対する前記第 2 の指向音声の非定常性度合いの非定常度比と前記パワー比の和に基づいて前記確からしさを算出する、付記 6 に記載の音声処理用コンピュータプログラム。

(付記 8)

集音した音声を表す第 1 の音声信号を生成する第 1 の音声入力部と、

前記第 1 の音声入力部と異なる位置に配置され、集音した音声を表す第 2 の音声信号を生成する第 2 の音声入力部と、

前記第 1 の音声信号及び第 2 の音声信号を、それぞれ、所定の時間長を持つフレームごとに周波数領域の第 1 の周波数スペクトル及び第 2 の周波数スペクトルに変換する時間周波数変換部と、

フレームごとに、前記第 1 の周波数スペクトル及び前記第 2 の周波数スペクトルに基づいて、受音することが優先される第 1 の方向及び前記第 1 の方向と異なる第 2 の方向のうちの前記第 2 の方向に位置する音源のみが音声を発した確からしさを算出する音源方向判定部と、

フレームごとに、前記第 1 の周波数スペクトル及び前記第 2 の周波数スペクトルに基づいて算出される前記第 1 の方向から到来する音声を含む第 1 の指向音声信号を出力するとともに、前記確からしさに応じて、前記第 1 の周波数スペクトル及び前記第 2 の周波数スペクトルに基づいて算出される前記第 2 の方向から到来する音声を含む第 2 の指向音声信号を出力するか否かを制御する指向音声出力部と、

10

20

30

40

50

を有する音声処理装置。

(付記9)

第1の音声入力部により生成された第1の音声信号、及び、前記第1の音声入力部と異なる位置に配置された第2の音声入力部により生成された第2の音声信号を、それぞれ、所定の時間長を持つフレームごとに周波数領域の第1の周波数スペクトル及び第2の周波数スペクトルに変換し、

フレームごとに、前記第1の周波数スペクトル及び前記第2の周波数スペクトルに基づいて、受音することが優先される第1の方向及び前記第1の方向と異なる第2の方向のうちの前記第2の方向に位置する音源のみが音声を発した確からしさを算出し、

フレームごとに、前記第1の周波数スペクトル及び前記第2の周波数スペクトルに基づいて算出される前記第1の方向から到来する音声を含む第1の指向音声信号を出力するとともに、前記確からしさに応じて、前記第1の周波数スペクトル及び前記第2の周波数スペクトルに基づいて算出される前記第2の方向から到来する音声を含む第2の指向音声信号を出力するか否かを制御する、

ことを含む音声処理方法。

【符号の説明】

【0084】

- 1 音声入力装置
- 11-1、11-2 マイクロホン
- 12-1、12-2 アナログ/デジタル変換器
- 13 音声処理装置
- 14 通信インターフェース部
- 21 時間周波数変換部
- 22 指向音声生成部
- 23 特徴抽出部
- 24 音源方向判定部
- 25 指向特性制御部
- 26 周波数時間変換部
- 100 コンピュータ
- 101 ユーザインターフェース部
- 102 オーディオインターフェース部
- 103 通信インターフェース部
- 104 記憶部
- 105 記憶媒体アクセス装置
- 106 プロセッサ
- 107 記憶媒体

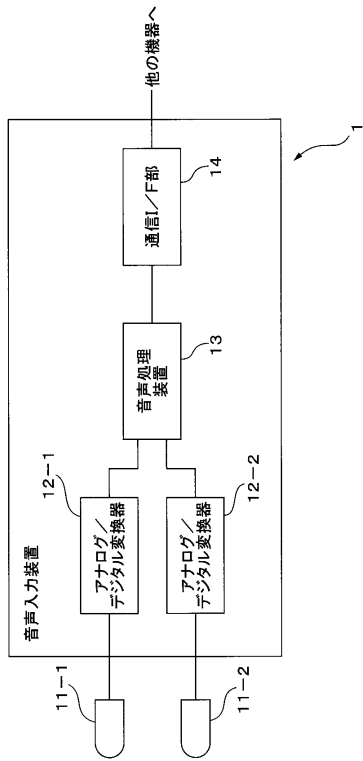
10

20

30

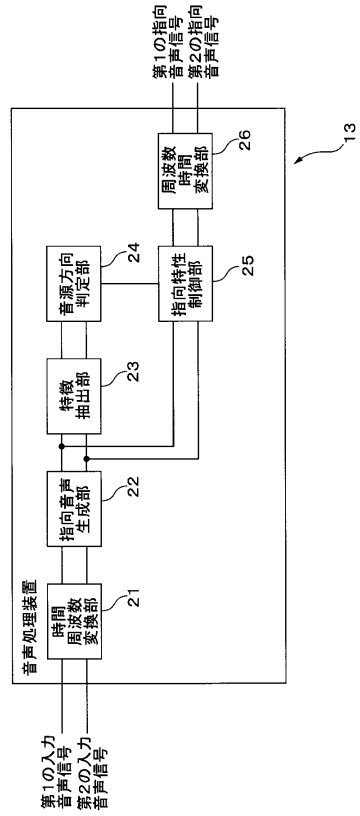
【 図 1 】

図1



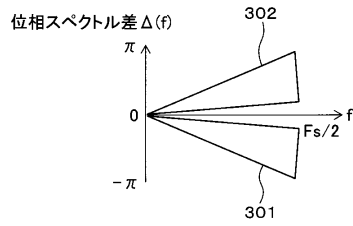
【 図 2 】

図2



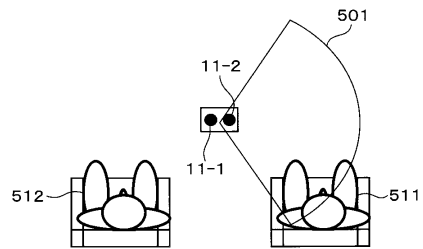
【 図 3 】

図3



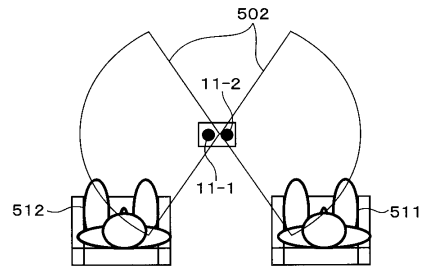
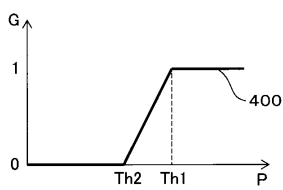
【 図 5 】

図5

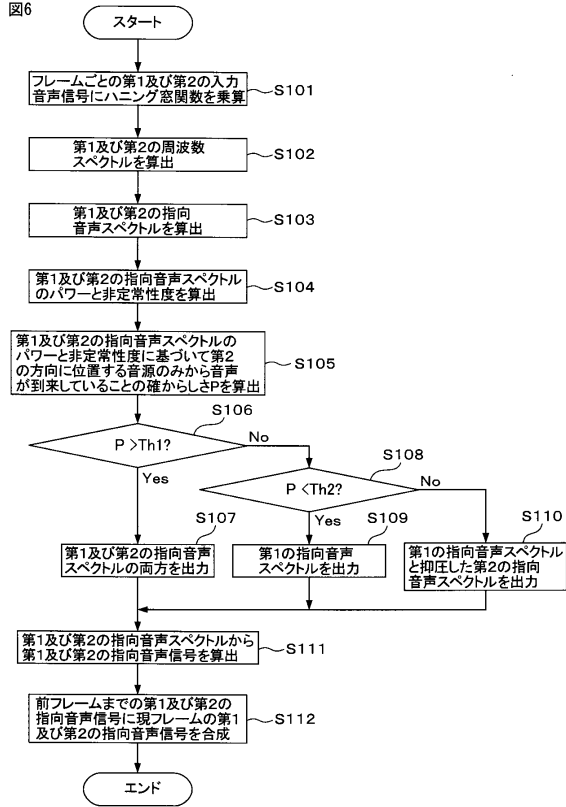


【 図 4 】

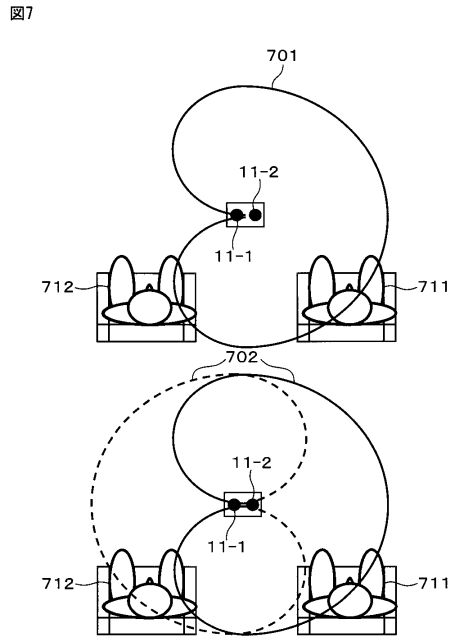
図4



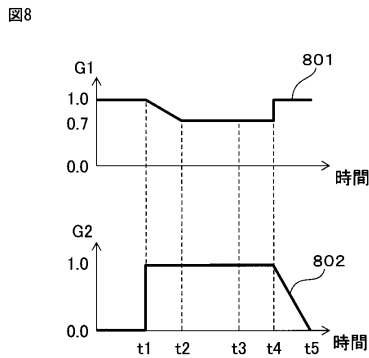
【図6】



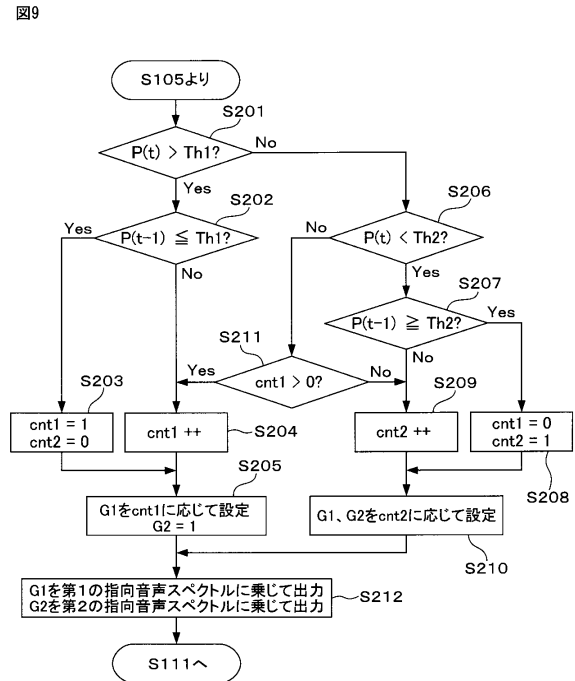
【図7】



【図8】

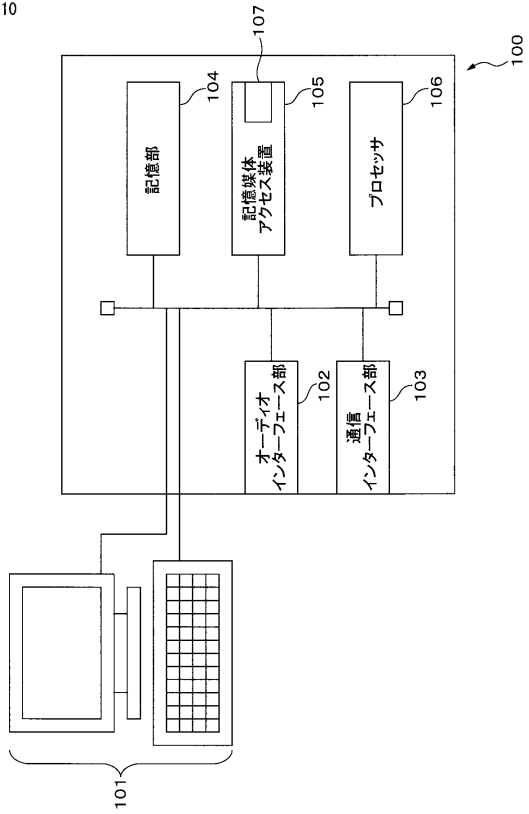


【図9】



【図10】

図10



フロントページの続き

審査官 上田 雄

- (56)参考文献 特開2017-9700(JP,A)
特開2006-58395(JP,A)
特開2007-219207(JP,A)
国際公開第2015/086895(WO,A1)

- (58)調査した分野(Int.Cl., DB名)
G10L 21/00-25/93