

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第3822397号

(P3822397)

(45) 発行日 平成18年9月20日(2006.9.20)

(24) 登録日 平成18年6月30日(2006.6.30)

(51) Int. Cl.

H03G 3/32 (2006.01)

F I

H03G 3/32

請求項の数 9 (全 15 頁)

(21) 出願番号	特願平11-272209	(73) 特許権者	000101732
(22) 出願日	平成11年9月27日(1999.9.27)		アルパイン株式会社
(65) 公開番号	特開2001-94370(P2001-94370A)		東京都品川区西五反田1丁目1番8号
(43) 公開日	平成13年4月6日(2001.4.6)	(74) 代理人	100103171
審査請求日	平成15年3月28日(2003.3.28)		弁理士 雨貝 正彦
		(72) 発明者	木内 真吾
			東京都品川区西五反田1丁目1番8号 ア
			ルパイン株式会社内
		(72) 発明者	斉藤 望
			東京都品川区西五反田1丁目1番8号 ア
			ルパイン株式会社内
		審査官	石井 研一

最終頁に続く

(54) 【発明の名称】 音声入出力方式

(57) 【特許請求の範囲】

【請求項1】

周辺ノイズ、オーディオ音、利用者による発声音声のそれぞれが存在する音響空間内の所定位置に設置されたスピーカおよび集音手段を備え、前記集音手段によって集音した利用者の発声音声に対して所定のゲイン補正を行って前記スピーカから前記音響空間内に放出する音声入出力方式において、

前記集音手段の出力信号の中から前記周辺ノイズに対応する成分を除去するノイズ除去手段と、

前記集音手段の出力信号の中から前記オーディオ音に対応する成分を除去するオーディオ音除去手段と、

前記集音手段の出力信号の中から、前記スピーカから放出されて前記集音手段に回り込む利用者自身の発声音声に対応する成分を除去する手段と、

前記集音手段の出力信号の中から、前記ノイズ除去手段、前記オーディオ音除去手段、前記回り込む利用者自身の発声音声に対応する成分を除去する手段のそれぞれによって前記周辺ノイズに対応する成分、前記オーディオ音に対応する成分、前記回り込む利用者自身の発声音声に対応する成分が除去された後の信号成分に対して、所定のゲイン補正を行う音声補正手段と、

前記音声補正手段によってゲイン補正が行われた後の信号成分を利用者の発声音声として前記スピーカから前記音響空間内に放出する音声出力手段と、

を備えることを特徴とする音声入出力方式。

【請求項 2】

請求項 1 において、

前記オーディオ音除去手段は、

前記音響空間の伝達特性に対応する第 1 のフィルタ係数を有し、前記オーディオ音に対応するオーディオ音信号が入力される第 1 のフィルタと、

前記集音手段の出力信号の中から、前記第 1 のフィルタを通した後の前記オーディオ音信号を差し引く第 1 の演算部と、

を備えることを特徴とする音声入出力方式。

【請求項 3】

請求項 2 において、

前記第 1 のフィルタは、適応等化処理を行う適応フィルタであり、前記第 1 の演算部から出力される差分信号のパワーが最小となるように前記第 1 のフィルタ係数が設定されることを特徴とする音声入出力方式。

【請求項 4】

請求項 3 において、

前記回り込む利用者自身の発生音声に対応する成分を除去する手段は、

前記音響空間の伝達特性に対応する第 2 のフィルタ係数を有し、前記スピーカから放出される利用者の発生音声に対応する信号が入力される第 2 のフィルタと、

前記集音手段の出力信号の中から、前記第 2 のフィルタを通した後の信号を差し引く第 2 の演算部と、

を備えることを特徴とする音声入出力方式。

【請求項 5】

請求項 4 において、

前記第 2 のフィルタ係数は、前記第 1 のフィルタ係数をコピーすることにより設定されることを特徴とする音声入出力方式。

【請求項 6】

請求項 1 ~ 5 のいずれかにおいて、

前記音声補正手段は、

前記周辺ノイズおよび前記オーディオ音の音圧レベルと、前記信号成分の音圧レベルとに基づいて、前記周辺ノイズおよび前記オーディオ音の音圧レベルによらず、前記スピーカから出力される発生音声の音圧レベルが静寂下と同じ大きさの音であると感じるために必要な補正ゲインを算出するゲイン算出手段と、

前記信号成分に対して前記ゲイン算出手段によって算出された前記補正ゲインに基づくゲイン補正を行うゲイン補正手段と、

を備えることを特徴とする音声入出力方式。

【請求項 7】

請求項 6 において、

前記ゲイン算出手段は、騒音下において静寂下と同じ大きさの音を感じるために前記発生音声の音圧レベルに対してどれだけゲインを加える必要があるかを示すゲインテーブルを様々な騒音レベル毎に有し、前記周辺ノイズおよび前記オーディオ音の音圧レベルとしての騒音レベルに対応する前記ゲインテーブルを用いて、前記発生音声の音圧レベルに対応する前記補正ゲインを算出することを特徴とする音声入出力方式。

【請求項 8】

請求項 6 または 7 において、

前記ゲイン算出手段は、複数の周波数成分毎に前記補正ゲインを算出し、

前記ゲイン補正手段は、前記ゲイン算出手段によって算出された複数の周波数成分毎の前記補正ゲインを用いてゲイン補正を行うことを特徴とする音声入出力方式。

【請求項 9】

請求項 1 ~ 8 のいずれかにおいて、

前記集音手段の出力信号の中から、前記ノイズ除去手段、前記オーディオ音除去手段、

10

20

30

40

50

前記回り込む利用者自身の発生音声に対応する成分を除去する手段のそれぞれによって前記周辺ノイズに対応する成分、前記オーディオ音に対応する成分、前記回り込む利用者自身の発生音声に対応する成分が除去された後の信号成分のみが入力される音声認識手段を備えることを特徴とする音声入出力方式。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、マイクロホンで集音した音声に所定の処理を加えて音声認識装置等に出力する音声入出力方式に関する。

【0002】

【従来の技術】

車両の走行案内を行うナビゲーション装置やオーディオ機器等の車載用機器においては、各種の操作指示を入力する方法として、利用者が操作パネルやリモートコントロールユニットに備えられた各種キーを押下する方法が従来から汎用されているが、最近では、利用者によって発せられた操作音声の内容を音声認識することによって操作指示入力を行う方法が用いられている。音声認識装置を用いて操作指示を行う場合には、操作キーの配置等を覚える必要がなく、しかも走行中に車両が振動した状態でキーの操作を行わないため、操作の簡略化が可能であり、最近では車載用機器に対する操作方法として用いられることが多くなっている。

【0003】

このような音声認識装置によって操作音声の内容を認識させる場合に、認識率を低下させる要因として代表的なものには、ロードノイズやエンジンノイズ等の走行に伴って生じる車室内の周辺ノイズと、車室内にオーディオ装置から出力されるオーディオ音とがある。これらの周辺ノイズやオーディオ音が、利用者が発声する操作音声に重畳した場合、音声認識装置ではこれらの入力音声から利用者の操作音声のみを区別して音声認識を行うことが困難となり、認識率が低下する。このため、従来は、適応マイクロホンアレイ技術を用いてロードノイズを低減させたり、トークスイッチが押下されたときにオーディオ音の出力を中断したり、あるいは音量を下げるなどして、音声認識の対象である利用者の操作音声に重畳される各種のノイズやオーディオ音を低減する工夫が行われている。

【0004】

【発明が解決しようとする課題】

ところで、上述した周辺ノイズやオーディオ音が大きい場合には、利用者が発声した操作音声は、マスキング効果によってかき消されてしまって利用者自身が自分の発声した操作音声を確認できない場合がある。このため、利用者は自分がどのように発声しているのかを認識できず、不安定な発音となって発声しにくくなるという問題がある。この場合には、当然ながら、音質や音量が不安定な音声が入力されることになるため、認識率の低下を招くという問題もある。

【0005】

上述したようにトークスイッチを押下してオーディオ音の出力を中断したり音量を下げることで、この問題点をある程度改善することができるが、ロードノイズ等が大きい場合もあるため、トークスイッチでは完全な対策とは言えない。また、トークスイッチを用いてオーディオ音の出力を中断したり音量を下げる場合には、操作音声の出力が頻繁になると、オーディオ音の出力が断続的になり、利用者によるオーディオ音の聴取を妨げるという新たな問題が生じる。特に、操作音声を発声することにより車載用機器の操作を行っている利用者以外の搭乗者においては、こうした操作とは無関係にオーディオ音を聴取している場合もあるため、聴取しているオーディオ音が頻繁に断続すると不快であり、認識対象となる入力音声以外の出力音を断続することなく操作音声のみを抽出することができる音声入出力方式が望まれている。

【0006】

本発明は、このような点に鑑みて創作されたものであり、その目的は、周囲のノイズ等が

10

20

30

40

50

大きい場合の利用者による発声のしにくさを改善することができる音声入出力方式を提供することにある。

【0007】

また、本発明の他の目的は、利用者の発声音声を抽出することにより、この音声に対して音声認識を行う際の認識率を向上させることができる音声入出力方式を提供することにある。

【0008】

【課題を解決するための手段】

上述した課題を解決するために、本発明の音声入出力方式は、周辺ノイズ、オーディオ音、利用者による発声音声のそれぞれが存在する音響空間内の所定位置にスピーカと集音手段を備え、集音手段によって集音した利用者の発生音声に対して所定のゲイン補正を行ってスピーカから音響空間内に放出しており、集音手段の出力信号の中から周辺ノイズに対応する成分を除去するノイズ除去手段と、集音手段の出力信号の中からオーディオ音に対応する成分を除去するオーディオ音除去手段と、集音手段の出力信号の中から、スピーカから放出されて集音手段に回り込む利用者自身の発生音声に対応する成分を除去する手段と、集音手段の出力信号の中から、ノイズ除去手段、オーディオ音除去手段、回り込む利用者自身の発生音声に対応する成分を除去する手段のそれぞれによって周辺ノイズに対応する成分、オーディオ音に対応する成分、回り込む利用者自身の発生音声に対応する成分が除去された後の信号成分に対して、所定のゲイン補正を行う音声補正手段と、音声補正手段によってゲイン補正が行われた後の信号成分を利用者の発生音声としてスピーカから音響空間内に放出する音声出力手段とを備えている。集音手段によって集音された信号の中から利用者の発声音声に対応した成分のみを抽出し、これにゲイン補正を行った後にスピーカから出力しており、利用者は、自分の発声内容をオーディオ音等の大きさにかかわらず常に確認することができるため、発声のしにくさを改善することができる。

また、上述したオーディオ音除去手段は、音響空間の伝達特性に対応する第1のフィルタ係数を有し、オーディオ音に対応するオーディオ音信号が入力される第1のフィルタと、集音手段の出力信号の中から、第1のフィルタを通した後のオーディオ音信号を差し引く第1の演算部とを備えることが望ましい。

また、上述した第1のフィルタは、適応等化処理を行う適応フィルタであり、第1の演算部から出力される差分信号のパワーが最小となるように第1のフィルタ係数が設定されることが望ましい。

また、上述した回り込む利用者自身の発生音声に対応する成分を除去する手段は、音響空間の伝達特性に対応する第2のフィルタ係数を有し、スピーカから放出される利用者の発生音声に対応する信号が入力される第2のフィルタと、集音手段の出力信号の中から、第2のフィルタを通した後の回り込む利用者自身の発生音声に対応する信号を差し引く第2の演算部とを備えることが望ましい。

また、上述した第2のフィルタ係数は、第1のフィルタ係数をコピーすることにより設定されることが望ましい。

【0009】

また、上述した音声補正手段は、周辺ノイズおよび前記オーディオ音の音圧レベルと、信号成分の音圧レベルとに基づいて、周辺ノイズおよびオーディオ音の音圧レベルによらず、スピーカから出力される発生音声が発音下と同じ大きさの音であると感じるために必要な補正ゲインを算出するゲイン算出手段と、信号成分に対してゲイン算出手段によって算出された補正ゲインに基づくゲイン補正を行うゲイン補正手段とを備えることが望ましい。

また、上述したゲイン算出手段は、騒音下において静寂下と同じ大きさの音を感じるために発生音声の音圧レベルに対してどれだけゲインを加える必要があるかを示すゲインテーブルを様々な騒音レベル毎に有し、周辺ノイズおよびオーディオ音の音圧レベルとしての騒音レベルに対応するゲインテーブルを用いて、発生音声の音圧レベルに対応する補正ゲインを算出することが望ましい。

10

20

30

40

50

また、上述したゲイン算出手段は、複数の周波数成分毎に補正ゲインを算出し、ゲイン補正手段は、ゲイン算出手段によって算出された複数の周波数成分毎の補正ゲインを用いてゲイン補正を行うことが望ましい。どの程度ゲインを補正した場合に明瞭に音声聞き取れるかは、全周波数領域で一律に決まるものではなく、周辺ノイズやオーディオ音あるいは発声音声の各周波数成分毎に異なるため、各周波数成分毎に補正ゲインを算出してゲイン補正を行うことにより、スピーカからより明瞭な音声を出力することができる。

【0010】

また、上述した集音手段の出力信号からこれらの各成分が除去された後の発声音声信号を用いて、音声認識手段による音声認識処理を行うことが望ましい。集音手段によって集音された音声にオーディオ音や周辺ノイズが含まれている場合であっても、利用者の発声音声のみを音声認識手段に入力することができるため、音声認識処理を行う際の認識率を高めることができる。また、利用者の発声のしにくさが改善されており、利用者は、安定した発声を行うことができるため、音声の調子等が発声の都度異なるといったことがなく、このような発声音声を用いて音声認識処理を行うことによってさらに認識率を高めることができる。

10

【0011】

【発明の実施の形態】

以下、本発明を適用した一実施形態の音声入出力装置について、図面を参照しながら説明する。

【0012】

〔第1の実施形態〕

図1は、本発明を適用した第1の実施形態の音声入出力装置の構成を示す図である。同図に示す音声入出力装置100は、マイクロホン110によって集音された各種の音声の中から利用者の発声音声のみを抽出して音声認識装置200に向けて出力するとともに、この発声音声に対してゲイン補正を行った後にスピーカ120から出力する。この音声入出力装置100は、適応フィルタ10、フィルタ12、演算部20、22、周辺ノイズ除去部30、ラウドネス補償演算部40、音声補正用フィルタ42、音声合成部50、アンプ52を含んで構成されている。

20

【0013】

適応フィルタ10は、車室内の音響空間の伝達特性を模擬するためのものであり、フィルタ係数(タップ係数)W1を有するFIR型のデジタルフィルタであって、オーディオ装置300から入力されるオーディオ音信号に対して所定の適応等化処理を行う。このフィルタ係数W1は、LMS(Least Mean Square)アルゴリズムによって、演算部22から出力される差分信号(後述する)のパワーが最小となるように更新される。フィルタ12は、適応フィルタ10と同様に車室内の音響空間の伝達特性を模擬するためのものであり、フィルタ係数W2を有している。フィルタ係数W2は、所定のタイミングで適応フィルタ10のフィルタ係数W1がコピーされる。

30

【0014】

演算部20は、マイクロホン110の出力信号とフィルタ12の出力信号とが入力され、これら2つの信号の差分を演算する。また、演算部22は、演算部20から出力される差分信号と適応フィルタ10の出力信号とが入力されており、これら2つの信号の差分を演算する。

40

【0015】

周辺ノイズ除去部30は、後段の演算部22から出力された差分信号に含まれる周辺ノイズに対応する成分を除去する。この周辺ノイズ除去部30からは、マイクロホン110から出力される信号に含まれる利用者の発声音声に対応する成分のみが抽出されて出力される。利用者の音声を抽出する詳細動作については後述する。

【0016】

ラウドネス補償演算部40は、オーディオ音信号および周辺ノイズ信号と利用者の発声音声信号とが入力されており、これらの信号に基づいて、利用者の発声音声をスピーカ12

50

0 から出力する際に必要な補正ゲインを算出する。音声補正用フィルタ 4 2 は、ラウドネス補償演算部 4 0 によって算出された補正ゲインに基づいて、周辺ノイズ除去部 3 0 から出力される音声信号に対するゲイン補正を行う。ラウドネス補償演算部 4 0 および音声補正用フィルタ 4 2 の詳細構成については後述する。

【 0 0 1 7 】

音声合成部 5 0 は、音声補正用フィルタ 4 2 によって所定のゲイン補正がなされた後の音声信号と、オーディオ装置 3 0 0 から入力されたオーディオ音信号とを合成する。音声合成部 5 0 から出力される合成信号は、アンプ 5 2 で増幅された後、スピーカ 1 2 0 から車室内に出力される。

【 0 0 1 8 】

上述したマイクロホン 1 1 0 が集音手段に、周辺ノイズ除去部 3 0 がノイズ除去手段に、適応フィルタ 1 0、演算部 2 2 がオーディオ音除去手段に、ラウドネス補償演算部 4 0、音声補正用フィルタ 4 2 が音声補正手段に、アンプ 5 2 が音声出力手段に、ラウドネス補償演算部 4 0 がゲイン算出手段に、音声補正用フィルタ 4 2 がゲイン補正手段に、音声認識装置 2 0 0 が音声認識手段にそれぞれ対応する。

【 0 0 1 9 】

本実施形態の音声入出力装置 1 0 0 はこのような構成を有しており、次にその動作を説明する。

【 0 0 2 0 】

オーディオ装置 3 0 0 から出力されたオーディオ音信号は、適応フィルタ 1 0 に入力されるとともに、音声合成部 5 0、アンプ 5 2 を介してスピーカ 1 2 0 から車室内に出力される。このスピーカ 1 2 0 から出力されたオーディオ音は、利用者の発声音声を聴取可能な所定位置に設定されたマイクロホン 1 1 0 によって集音されるため、マイクロホン 1 1 0 から出力されて演算部 2 0 を介して演算部 2 2 の一方の入力端に入力される信号にはオーディオ音に対応する成分が含まれている。また、このオーディオ音は、車室内に出力された後にマイクロホン 1 1 0 で集音されたものであるため、車室内の音響空間の伝達特性が反映されたものである。

【 0 0 2 1 】

したがって、車室内の音響空間の伝達特性が反映されたオーディオ音に対応する成分が含まれるマイクロホン 1 1 0 の出力信号と、オーディオ装置 3 0 0 から直接入力されたオーディオ音信号を適応フィルタ 1 0 に通した後の信号との差分を演算部 2 2 で演算し、この差分信号のパワーが最小となるように適応フィルタ 1 0 のフィルタ係数 W_1 を更新することにより、このフィルタ係数 W_1 は車室内の音響空間の伝達特性を模擬したものとなる。すなわち、演算部 2 2 の一方の入力端に入力される信号には、オーディオ装置 3 0 0 から出力されて実際の車室内の音響空間に出力されたオーディオ音に対応する成分が含まれており、他方の入力端に入力される信号には、この音響空間の特性を模擬した適応フィルタ 1 0 を通した後のオーディオ音に対応する成分が含まれることになり、演算部 2 2 によってこれらの差分を演算することにより、オーディオ音に対応する成分が除去される。また、演算部 2 2 の後段には周辺ノイズ除去部 3 0 が配置されており、演算部 2 2 の出力信号に含まれる周辺ノイズが除去される。

【 0 0 2 2 】

このように、適応フィルタ 1 0 と演算部 2 2 によってオーディオ音に対応する成分が除去され、さらに周辺ノイズ除去部 3 0 によって周辺ノイズに対応する成分が除去される。したがって、利用者の発声音声と、スピーカ 1 2 0 から出力されるオーディオ音と、走行雑音やエンジン雑音等の周辺ノイズとが同時にマイクロホン 1 1 0 によって集音された場合であっても、これらが重畳されたマイクロホン 1 1 0 の出力信号の中からオーディオ音と周辺ノイズに対応する成分が除去され、周辺ノイズ除去部 3 0 からは、利用者の発声音声に対応する成分のみを出力することができる。このため、音声認識装置 2 0 0 では、利用者の発声音声のみに対して音声認識処理を行うことができ、認識率を高めることができる。

【0023】

また、本実施形態の音声入出力装置100は、利用者の周囲がオーディオ音や周辺ノイズの存在によって騒がしい場合であって、自分が発声した音声を直接聴取できないために、発声が不安定になることを防止するために、利用者の音声を拡声してスピーカ120から出力する機能を有しており、次にその詳細について説明する。

【0024】

〔ラウドネス補償演算部の詳細〕

ラウドネス補償演算部40は、次に説明する原理に基づいて、スピーカ120から出力する利用者の発声音声に対して各周波数成分の信号レベルを調整するために必要な最適なゲインを算出する。

10

【0025】

図2は、物理的な音圧レベルと、その音を人間が聞いたときに感じる音の大きさ（ラウドネス）との対応関係（ラウドネス曲線）を示す図である。同図において、横軸は音圧レベル（単位：dB - SPL）、縦軸は人間が感じる音の大きさを示すラウドネス（単位：sone）であり、曲線1は静寂下でのラウドネス曲線、曲線2は騒音下でのラウドネス曲線である。ただし、曲線2は騒音レベルに応じて変化するものである。

【0026】

同図において、ラウドネスの値が同じであれば、人間は同じ大きさの音であると感じる。したがって、例えば、人間が0.1soneの大きさに感じる音は、静寂下では約12dB - SPLの音圧レベルであるが、曲線2に示す騒音下では約37dB - SPLの音圧レベルの音である。すなわち静寂下で約12dB - SPLで出力していた音を曲線2の騒音下で同じ大きさに感じるには約37dB - SPLの音を出力する必要があり、約25dBのゲインを加える必要があるということである。また、人間が1soneの大きさに感じる音は、静寂下では約42dB - SPLの音圧レベルの音であるが、曲線2の騒音下では約49dB - SPLの音圧レベルであるため、騒音下では約7dBのゲインを加えてやる必要がある。したがって、同じ騒音下でも、出力される音の音圧レベルに応じて加えるゲインを変更する必要があるということである。

20

【0027】

図3は、騒音下において静寂下と同じ大きさの音を感じるために、静寂下の音圧レベルに対してどれだけゲインを加える必要があるかを示す図である。同図において、横軸は静寂下で出力される音の音圧レベルであり、縦軸は騒音下において静寂下と同じ大きさの音を感じるために加える必要があるゲイン値である。例えば、静寂下で音圧レベル20dBで出力される音は、騒音下では、約19dBのゲインを加えられることによって、人間は静寂下と同じ大きさの音であると感じるようになる。

30

【0028】

ラウドネス補償演算部40は、あらかじめ様々な騒音レベルにおける図3に示すような音声信号の音圧レベル（周辺ノイズ除去部30から出力される利用者の発声による音声の音圧レベル）と加えるゲインとの関係（以下、ゲインテーブルと呼ぶ）を内部のメモリに格納しており、入力されるオーディオ音信号と周辺雑音信号に基づいて、最適なゲインテーブルを選択し、この選択したゲインテーブルと周辺ノイズ除去部14から出力される音声信号とに基づいて、最適なゲインを算出する。ラウドネス補償演算部40は、この算出されたゲインを音声補正用フィルタ42に出力して音声信号に対して最適なゲインを与える。

40

【0029】

ところで、一般にオーディオ音や周辺雑音は、様々な周波数成分を有しており、その周波数成分ごとに音圧レベルが異なっている。したがって、利用者が発声した音声をスピーカ120から出力しようとした場合に、この音声の聴き取りやすさが出力音声の音圧レベルだけでなく、オーディオ音や周辺雑音の各周波数成分の音圧レベルによっても異なるという不均衡が生じる。また、オーディオ音や周辺雑音の各周波数成分はそれらの高周波成分の発聲音声に対してマスキング効果を及ぼすため、このことも考慮する必要がある。

50

【 0 0 3 0 】

そこで、音声信号の各周波数成分ごとに最適なゲインを与えることが望ましい。すなわち、音声信号とオーディオ音信号および周辺雑音信号のそれぞれを所定の周波数帯域に分割して、各周波数帯域ごとにオーディオ音信号・周辺雑音信号の周波数成分に基づいて最適なゲインテーブルを選択し、この選択したゲインテーブルと音声信号の周波数成分とに基づいて最適なゲインを算出することが望ましい。

【 0 0 3 1 】

図4は、ラウドネス補償演算部40の詳細構成を示す図である。同図に示すようにラウドネス補償演算部40は、周波数帯域レベル平均部410、ラウドネス算出部412、周波数帯域ゲインテーブル選択部414、周波数帯域レベル平均部418、ゲインテーブル416を含んで構成されている。

10

【 0 0 3 2 】

周波数帯域レベル平均部410は、適応フィルタ10から入力されるオーディオ音信号と演算部22から入力される周辺ノイズ信号(以下、雑音等と呼ぶ)に対して、所定の時間ブロックごとに周知のFFT(Fast Fourier Transform)演算を行い、所定の周波数帯域ごとに音圧レベルの平均を計算する。雑音等は、例えば人間の聴覚がほぼ1/3オクターブごとに音の大きさの違いを認識することができるという特性を考慮して1/3オクターブごとに周波数分割される。

【 0 0 3 3 】

なお、マイクロホン110に向かって利用者が発声すると、演算部22から出力される差分信号にはこの利用者の発声に対応する成分も含まれるため、利用者が発声を開始する直前に演算部22から出力される周辺ノイズ信号のみをラウドネス補償演算部40において取り込むようにする。例えば、トークスイッチを設けておいて、利用者に発声する直前にこのトークスイッチを押下させるようにすればよい。

20

【 0 0 3 4 】

ラウドネス算出部412は、周知のZwickerのラウドネス算出手法(ISO 532B)やStevensのラウドネス算出手法(ISO 532A)を用いて、周波数帯域レベル平均部410から周波数帯域ごとに出力される雑音等の音圧レベルを調整する。具体的には、以下のように調整を行う。すなわち、ある周波数成分の雑音等があるとき、この雑音等は、同一の周波数成分の発声音声の聴き取りにくさに影響するのみならず、マスキング効果により高周波側に隣接する周波数成分の発声音声の聴き取りにくさにも影響を与える。ラウドネス算出部412は、これを考慮して、雑音等の各周波数成分の音圧レベルを低周波側に隣接する雑音等の周波数成分の音圧レベルの大きさに応じて調整を行う。すなわち、隣接する低周波成分の音圧レベルが大きい場合には、高周波側に隣接する周波数成分の音圧レベルを高め補正する。このような調整を行うことで、各周波数帯域毎のゲインテーブルを選択する際には、対応する各周波数帯域の雑音等の音圧レベルに着目するのみで足り、低周波側に隣接する周波数帯域の雑音等を考慮するという煩雑な処理を行う必要がなくなる。

30

【 0 0 3 5 】

周波数帯域ゲインテーブル選択部414は、ラウドネス算出部412から出力される調整後の周波数帯域ごとの雑音等の音圧レベルに基づいて、周波数帯域ごとに最適なゲインテーブル416を選択する。

40

【 0 0 3 6 】

周波数帯域レベル平均部418は、周辺ノイズ除去部30から入力される発聲音声信号に対して、短時間のブロックごとに周知のFFT演算を行い、所定の周波数帯域ごとに音圧レベルの平均を計算する。発聲音声信号は、雑音等と同様の周波数帯域に分割される。周波数帯域レベル平均部418から出力される周波数帯域ごとに分割された発聲音声信号は、周波数帯域ゲインテーブル選択部414によって選択されたゲインテーブル416に入力され、各周波数帯域ごとに適切なゲイン値が算出される。

【 0 0 3 7 】

50

このように、雑音等や発声音声信号を所定の周波数帯域に分割することによって、各周波数帯域ごとにゲインテーブルを選択して発声音声信号に最適なゲインを加えることが可能となる。

【 0 0 3 8 】

上述したラウドネス補償演算部 4 0 では、周波数帯域レベル平均部 4 1 0 および 4 1 8 を用いて発声音声信号や雑音等の周波数帯域ごとの音圧レベルの平均を求めたが、これらの周波数帯域レベル平均部の代わりにフィルタバンクとブロック平均部を用いて周波数帯域毎の音圧レベルの平均を求めるようにしてもよい。

【 0 0 3 9 】

〔音声補正用フィルタの詳細〕

次に、音声補正用フィルタ 4 2 の詳細について説明する。音声補正用フィルタ 4 2 は、上述したラウドネス補償演算部 4 0 で算出されたゲイン特性を修正（ゲインの加算）できるものであればよい。様々な構成が考えられるが、その一例として以下の 3 通りの構成について説明する。

【 0 0 4 0 】

図 5 は、フィルタバンクと可変ゲイン部を用いた音声補正用フィルタ 4 2 の構成を示す図である。同図に示す音声補正用フィルタ 4 2 は、フィルタバンク 4 2 0、可変ゲイン部 4 2 2、加算器 4 2 4 を含んで構成されている。

【 0 0 4 1 】

フィルタバンク 4 2 0 は、所定の周波数帯域幅を持つバンドパスフィルタ群であり、これらのバンドパスフィルタ群によって発声音声信号を周波数帯域ごとに分割する。可変ゲイン部 4 2 2 は、ラウドネス補償演算部 4 0 によって算出された各周波数帯域ごとのゲインを、フィルタバンク 4 2 0 から出力される周波数帯域ごとに分割された発声音声信号の音圧レベルに与えて、ゲイン調整を行う。加算器 4 2 4 は、各周波数帯域ごとにゲイン調整された発声音声信号を足し合わせて出力して、所望のゲイン補正を実現する。この構成によれば、アナログ回路で安価に音声補正用フィルタ 4 2 を構成することができる。

【 0 0 4 2 】

図 6 は、周波数サンプリングフィルタを用いた音声補正用フィルタ 4 2 の構成を示す図である。同図に示す音声補正用フィルタ 4 2 は、スプライン関数補間部 4 3 0、IFFT 演算部 4 3 2、FIR フィルタ 4 3 4 を含んで構成されている。

【 0 0 4 3 】

スプライン関数補間部 4 3 0 は、ラウドネス補償演算部 4 0 によって算出された各周波数帯域のゲインをそれぞれの周波数帯域の中心周波数のゲインとして、それぞれのゲイン値の間を周知のスプライン関数を用いて補間することによって周波数領域における滑らかなゲイン特性を得る。IFFT 演算部 4 3 2 は、スプライン関数補間部 4 3 0 から出力されるゲイン特性を周知の IFFT (Inverse Fast Fourier Transform) 演算を用いて周波数領域から時間領域に変換し、FIR フィルタ 4 3 4 のタップ係数の値を設定する。FIR フィルタ 4 3 4 は、発声音声信号に対して時間軸上のフィルタリング処理を行い、所望のゲイン補正を実現する。この構成によれば、直線位相フィルタを実現することができ、発声音声信号に対する補正は、周波数帯域ごとではなく、周波数成分ごとに行うことが可能となる。

【 0 0 4 4 】

図 7 は、周波数領域フィルタを用いた音声補正用フィルタ 4 2 の構成を示す図である。同図に示す音声補正用フィルタ 4 2 は、スプライン関数補間部 4 4 0、FFT 演算部 4 4 2、周波数帯域フィルタリング部 4 4 4、IFFT 演算部 4 4 6 を含んで構成されている。

【 0 0 4 5 】

スプライン関数補間部 4 4 0 は、ラウドネス補償演算部 4 0 によって算出された各周波数帯域のゲインをそれぞれの周波数帯域の中心周波数のゲインとして、それぞれのゲイン値の間を周知のスプライン関数を用いて補間することによって周波数領域における滑らかなゲイン特性を得る。FFT 演算部 4 4 2 は、発声音声信号に対して FFT 演算を行い、時

10

20

30

40

50

間領域から周波数領域に変換する。周波数帯域フィルタリング部444は、FFT演算部442から出力される周波数領域における発声音声信号に対して、スプライン関数補間部440から出力される滑らかなゲイン特性によってフィルタリングを行い、IFFT演算部446は、周波数帯域フィルタリング部444から出力される周波数領域における発声音声に対してIFFT演算を行って周波数領域から時間領域に変換して、所望のゲイン補正を実現する。IFFT演算の過程においては、線形フィルタリングを実現するために周知の重畳加算法(overlap-add method)や重畳保留法(overlap-save method)を用いるとよい。この構成によって、フィルタのタップ数が多いときでも演算量を比較的少なくすることができる。

【0046】

なお、上述した3通りの音声補正フィルタ42においては、いずれの場合もゲインが急激に変化すると出力波形が不連続になってしまうため、

$$G(n) = G(n-1) + G_m$$

を用いて、ゲイン特性を徐々に更新することが好ましい。ここで、 $G(n)$ は時間 n におけるゲイン特性、 $G(n-1)$ は時間 $n-1$ におけるゲイン特性、 G_m はラウドネス補償演算部40やスプライン関数補間部430、440によって算出されたゲイン特性である。 G_m は係数で $G_m + G(n-1) = 1$ になる関係がある。

【0047】

このように、本実施形態の音声入出力装置100では、ラウドネス補償算出部40および音声補正用フィルタ42を用いることにより、周辺ノイズ除去部30から出力される発声音声信号をスピーカ120から出力した際に、同じ車室内の音響空間に出力されたオーディオ音や周辺ノイズの音圧レベルに関係なく、発声音声が常に良好に聴取可能なように各周波数帯域のゲイン調整が行われる。したがって、利用者は、自分の発声内容を確認しながら発声を継続することができるため、発声のしにくさを改善することができる。このため、常に安定した状態で各種の操作音声等を発声することができ、音声認識装置200に入力される音声信号の状態も安定するようになるため、さらに音声認識処理の認識率を高めることができる。

【0048】

〔第2の実施形態〕

図8は、本発明を適用した第2の実施形態の音声入出力装置の構成を示す図である。なお、本実施形態の音声入出力装置100Aの構成において、図1に示した第1の実施形態の音声入出力装置100の構成と同じ動作を行うものについては同じ符号を付し、詳細な説明は省略する。

【0049】

図8に示す本実施形態の音声入出力装置100Aは、図1に示した第1の実施形態の音声入出力装置100の機能に加えて、ナビゲーション装置(図示せず)等から出力された案内音声の明瞭度を増す補正を行う機能を有する。この音声入出力装置100Aは、適応フィルタ10、フィルタ12、演算部20、22、周辺ノイズ除去部30、ラウドネス補償演算部40、音声補正用フィルタ42、44、音声合成部50、アンプ52、トークスイッチ60、スイッチ70、72、74、76を含んで構成されている。

【0050】

トークスイッチ60は、上述した2つの機能を切り替えるために、利用者自身によって操作される。例えば、利用者が何らかの操作音声を発声しようとしてトークスイッチ60を操作すると、この操作に応じた切替信号が4つのスイッチ70~76に送られる。

【0051】

スイッチ70、72は、2つの入力端子のそれぞれに入力される信号を、トークスイッチ60から入力される切替信号の有無に応じて選択的に出力する。具体的には、スイッチ70の一方の入力端子には周辺ノイズ除去部30の出力信号が入力され、他方の入力端子にはナビゲーション装置(図示せず)等から出力される案内音声信号が入力されている。トークスイッチ60が操作されて切替信号が出力されると、一方の入力端子側の接続状態が

10

20

30

40

50

有効になり、以後周辺ノイズ除去部30から出力される信号がスイッチ70を介してフィルタ12および音声補正用フィルタ42に入力される。また、トークスイッチ60が操作されない状態においては、他方の入力端子側の接続状態が有効になり、ナビゲーション装置等から入力される案内音声信号がスイッチ70を介してフィルタ12および音声補正用フィルタ42に入力される。なお、図1に示した音声入出力装置100に比べてフィルタ12の配置が異なっているが、基本的な動作に違いはなく、このフィルタ12によって、スピーカ120から出力されてマイクロホン110に回り込んで集音される発声音声のエコー成分が除去される。

【0052】

また、スイッチ72の一方の入力端子には周辺ノイズ除去部30の出力信号が入力され、他方の入力端子には音声補正用フィルタ44の出力信号が入力されている。トークスイッチ60が操作されて切替信号が出力されると、一方の入力端子側の接続状態が有効になり、以後周辺ノイズ除去部30から出力された信号がスイッチ72を介してラウドネス補償演算部40に入力される。また、トークスイッチ60が操作されない状態においては、他方の入力端子の接続状態が有効になり、音声補正用フィルタ44の出力信号がスイッチ72を介してラウドネス補償演算部40に入力される。なお、音声補正用フィルタ44は、ラウドネス補償演算部40によってゲインが設定された音声補正用フィルタ42の特性をコピーしたものである。

【0053】

また、スイッチ74、76は、トークスイッチ60から出力される切替信号の有無に応じて、オン状態とオフ状態が切り替えられる。スイッチ74は、トークスイッチ60が操作されて切替信号が出力されるとオン状態になり、適応フィルタ10から出力される信号を演算部22およびラウドネス補償演算部40に向けて出力する。また、スイッチ76は、トークスイッチ60が操作されず、切替信号が出力されないときにオン状態になり、演算部22から出力される信号をラウドネス補償演算部40に向けて出力する。

【0054】

トークスイッチ60が操作されて切替信号が出力された場合の各スイッチ70～76の接続状態は、上述した第1の実施形態の音声入出力装置100と基本的に同じであり、マイクロホン110の出力信号に含まれるオーディオ音に対応する成分と、周辺ノイズに対応する成分とが除去されて、利用者の発声音声に対応する成分のみが音声認識装置200に向けて出力される。また、この利用者の発声音声は、音声補正用フィルタ42を通すことにより所定のゲイン補正が行われた後にアンプ52によって増幅され、スピーカ120から出力されるため、利用者は、自分の発声内容を確認しながら発声を継続することができ、発声のしにくさを改善することができる。

【0055】

なお、トークスイッチ60が操作されると、スイッチ76がオフ状態になって、演算部22から出力される信号(周辺ノイズ信号)がラウドネス補償演算部40に入力されないことになるが、ラウドネス補償演算部40では、スイッチ76がオフ状態になる直前に入力された周辺ノイズ信号を用いてその後のゲイン算出を行っている。特に、周辺ノイズについては、短時間でのパワーの変動が少ないと考えられるため、このようにしても実用上支障はない。

【0056】

また、トークスイッチ60が操作されない状態においては、ナビゲーション装置等から入力された案内音声信号がスイッチ70、フィルタ12、音声補正用フィルタ44、スイッチ72を介してラウドネス補償演算部40に入力されるとともに、演算部22から出力される周辺ノイズ信号およびオーディオ音信号がスイッチ76を介してラウドネス補償演算部40にそれぞれ入力される。ラウドネス補償演算部40は、入力されるそれぞれの信号に基づいて音声補正用フィルタ42のゲインを設定する。したがって、ナビゲーション装置等から入力された案内音声は、スピーカ120から出力した際に、同じ車室内の音響空間に出力されたオーディオ音や周辺ノイズの音圧レベルに関係なく、常に良好に聴取可能

10

20

30

40

50

なように各周波数帯域のゲイン調整が行われる。このため、利用者は、オーディオ音や周辺ノイズが大きい場合であっても、スピーカ 120 から出力される案内音声の内容を明瞭に聴取することができる。

【0057】

なお、本発明は上記実施形態に限定されるものではなく、本発明の要旨の範囲内において種々の変形実施が可能である。例えば、上述した実施形態では、車載用の音声入出力装置について説明したが、音声入出力装置の用途は車載用に限定されず、建物内あるいは屋外で用いるようにしてもよい。

【0058】

【発明の効果】

上述したように、本発明によれば、集音手段によって集音された信号の中から利用者の発声音声に対応した成分のみを抽出し、これにゲイン補正を行った後にスピーカから出力しており、利用者は、自分の発声内容をオーディオ音等の大きさにかかわらず常に確認することができるため、発声のしにくさを改善することができる。

【0059】

また、本発明によれば、集音された音声にオーディオ音や周辺ノイズが含まれている場合であっても、利用者の発声音声のみを抽出することができるため音声認識処理を行う際の認識率を高めることができる。特に、利用者の発声のしにくさが改善されており、利用者は、安定した発声を行うことができるため、音声の調子等が発声の都度異なるといったことがなく、このような発聲音声を用いて音声認識処理を行うことによってさらに認識率を

【図面の簡単な説明】

【図1】第1の実施形態の音声入出力装置の構成を示す図である。

【図2】音圧レベルとその音を人間が聞いたときに感じる音の大きさとの対応関係を示す図である。

【図3】騒音下において静寂下と同じ大きさの音に感じるために、静寂下の音圧レベルに対してどれだけゲインを加える必要があるかを示す図である。

【図4】ラウドネス補償演算部の詳細構成を示す図である。

【図5】フィルタバンクと可変ゲインを用いた音声補正用フィルタの構成を示す図である。

【図6】周波数サンプリングフィルタを用いた音声補正用フィルタの構成を示す図である。

【図7】周波数領域フィルタを用いた音声補正用フィルタの構成を示す図である。

【図8】第2の実施形態の音声入出力装置の構成を示す図である。

【符号の説明】

- 10 適応フィルタ
- 12 フィルタ
- 20、22 演算部
- 30 周辺ノイズ除去部
- 40 ラウドネス補償演算部
- 42、44 音声補正用フィルタ
- 50 音声合成部
- 52 アンプ
- 60 トークスイッチ
- 70、72、74、76 スイッチ
- 100、100A 音声入出力装置
- 110 マイクロホン
- 120 スピーカ
- 200 音声認識装置
- 200 オーディオ装置

10

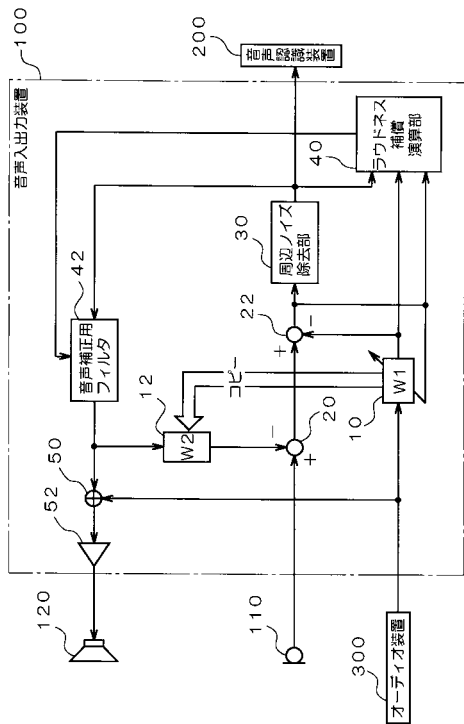
20

30

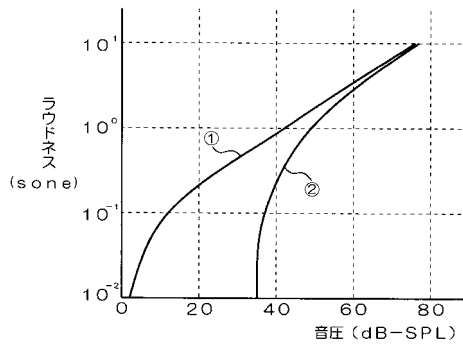
40

50

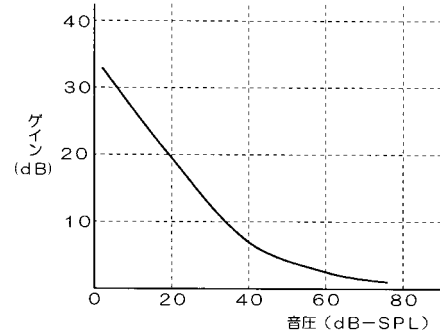
【 図 1 】



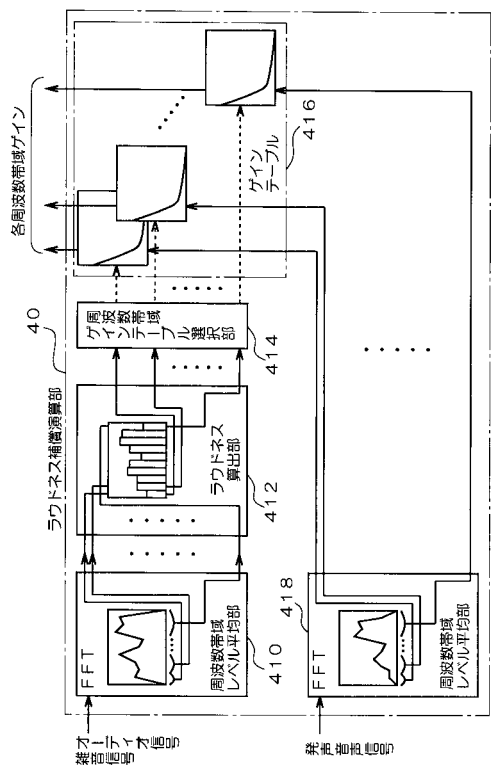
【 図 2 】



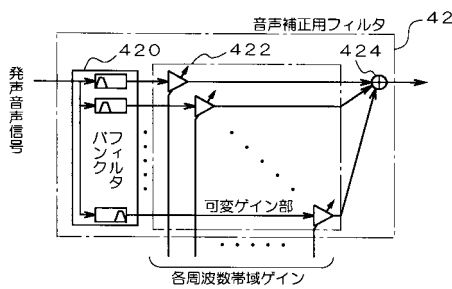
【 図 3 】



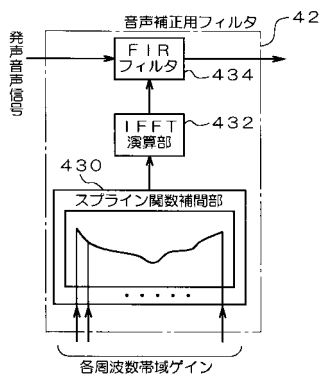
【 図 4 】



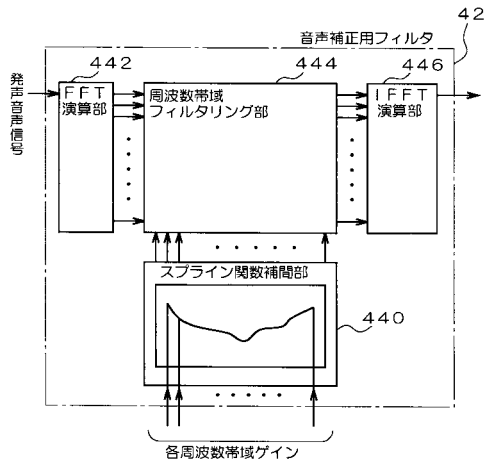
【 図 5 】



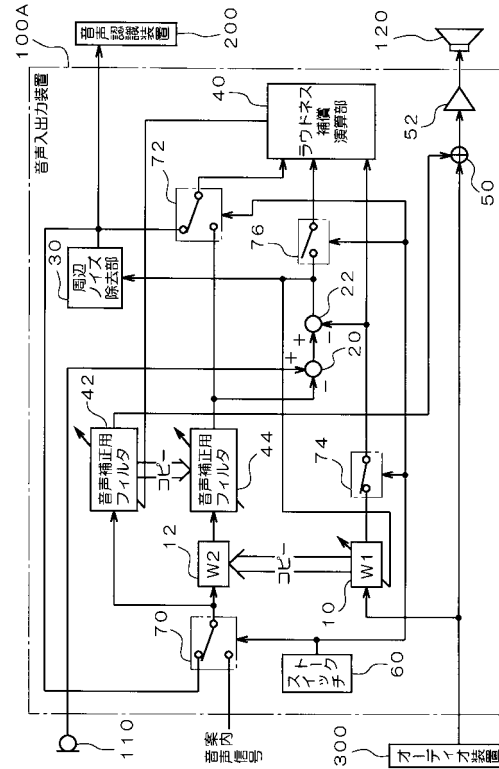
【 図 6 】



【 図 7 】



【 図 8 】



フロントページの続き

- (56)参考文献 特開平02 - 244099 (JP, A)
特開平04 - 235600 (JP, A)
特開平11 - 166835 (JP, A)
特開平05 - 259779 (JP, A)
特開2000 - 231399 (JP, A)
特開昭58 - 205397 (JP, A)
特開昭60 - 239199 (JP, A)

- (58)調査した分野(Int.Cl., DB名)

H03G 3/32