

(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2016-127300
(P2016-127300A)

(43) 公開日 平成28年7月11日(2016.7.11)

(51) Int.Cl.	F 1	テーマコード (参考)
H04R 3/00 (2006.01)	H04R 3/00 320	5D220
G10L 15/28 (2013.01)	G10L 15/28 400	
G10L 15/20 (2006.01)	G10L 15/20 370E	
G10L 15/00 (2013.01)	G10L 15/20 370D	
	G10L 15/20 370F	

審査請求 未請求 請求項の数 6 OL (全 16 頁) 最終頁に続く

(21) 出願番号 特願2014-263918 (P2014-263918)
(22) 出願日 平成26年12月26日 (2014.12.26)

(71) 出願人 000000011
アイシン精機株式会社
愛知県刈谷市朝日町2丁目1番地
(74) 代理人 100094112
弁理士 岡部 譲
(74) 代理人 100106183
弁理士 吉澤 弘司
(74) 代理人 100170601
弁理士 川崎 孝
(72) 発明者 ヴラジック サシヤ
フランス共和国 75002 パリ, リュ
ヴィヴィエヌ 12, フリーインパル
ス内
Fターム(参考) 5D220 BA04 BA06 BB04 BC05

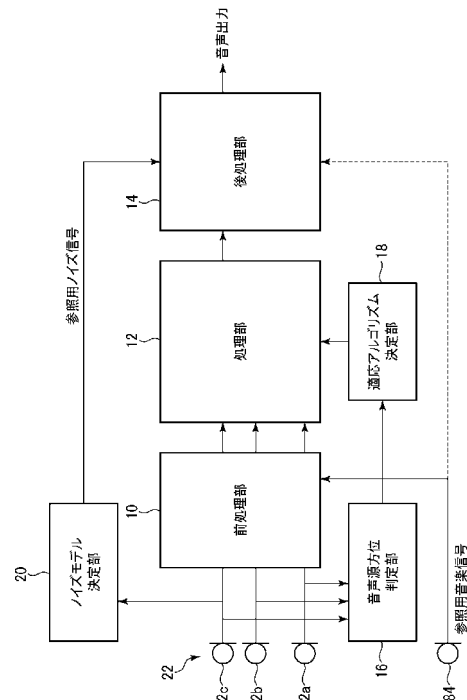
(54) 【発明の名称】 音声処理装置

(57) 【要約】

【課題】 音声認識の确实性を向上し得る音声処理装置を提供する。

【解決手段】 車両に配された複数のマイクロフォン22と、複数のマイクロフォンの各々によって取得される受音信号に含まれる音声の発生源である音声源が近傍界に位置する場合には、受音信号を球面波として扱って音声源の方位を判定し、音声源が遠方界に位置する場合には、受音信号を平面波として扱って音声源の方位を判定する音声源方位判定部16と、音声源の方位を含む方位範囲以外の方位範囲から到来する音を抑圧するようにビームフォーミングを行うビームフォーミング処理部12とを有している。

【選択図】 図2



【特許請求の範囲】**【請求項 1】**

車両に配された複数のマイクロフォンと、

前記複数のマイクロフォンの各々によって取得される受信信号に含まれる音声の発生源である音声が近傍界に位置する場合には、前記受信信号を球面波として扱って前記音源の方位を判定し、前記音声が前記遠方界に位置する場合には、前記受信信号を平面波として扱って前記音源の方位を判定する音声源方位判定部と、

前記音源の前記方位を含む方位範囲以外の方位範囲から到来する音を抑圧するようにビームフォーミングを行うビームフォーミング処理部と
を有することを特徴とする音声処理装置。

10

【請求項 2】

前記複数のマイクロフォンの数は、2つである
ことを特徴とする請求項 1 記載の音声処理装置。

【請求項 3】

前記複数のマイクロフォンの数は、少なくとも3つであり、

前記複数のマイクロフォンのうちの第1のマイクロフォンと前記複数のマイクロフォンのうちの第2のマイクロフォンとの間の距離である第1の距離は、前記複数のマイクロフォンのうちの第3のマイクロフォンと前記第2のマイクロフォンとの間の距離である第2の距離とは異なる

ことを特徴とする請求項 1 記載の音声処理装置。

20

【請求項 4】

前記受信信号に混入された音楽信号を、音響機器から取得された参照用音楽信号を用いて除去する音楽除去処理部を更に有する

ことを特徴とする請求項 1 乃至 3 のいずれか 1 項に記載の音声処理装置。

【請求項 5】

前記音声源方位判定部は、前記第2の方位範囲内から前記マイクロフォンに到来した音が、前記第1の方位範囲内から前記マイクロフォンに到来した音よりも大きい場合には、前記音源の前記方位の判定を中断する

ことを特徴とする請求項 1 乃至 4 のいずれか 1 項に記載の音声処理装置。

【請求項 6】

前記受信信号に混入されたノイズの除去処理を行うノイズ除去処理部を更に有する

ことを特徴とする請求項 1 乃至 5 のいずれか 1 項に記載の音声処理装置。

30

【発明の詳細な説明】**【技術分野】****【0001】**

本発明は、音声処理装置に関する。

【背景技術】**【0002】**

自動車等の車両には、様々な機器が設けられている。これらの様々な機器に対する操作は、例えば、操作ボタンや操作パネル等を操作することにより行われている。

40

【0003】

一方、近時では、音声認識の技術も提案されている（特許文献 1～3）。

【先行技術文献】**【特許文献】****【0004】**

【特許文献 1】特開 2012 - 215606 号公報

【特許文献 2】特開 2012 - 189906 号公報

【特許文献 3】特開 2012 - 42465 号公報

【発明の概要】**【発明が解決しようとする課題】**

50

【0005】

しかしながら、車両においては、様々なノイズが存在する。このため、車両内で発せられる音声に対しての音声認識は容易ではなかった。

【0006】

本発明の目的は、音声認識の確実性を向上し得る良好な音声処理装置を提供することにある。

【課題を解決するための手段】

【0007】

本発明の一観点によれば、車両に配された複数のマイクロフォンと、前記複数のマイクロフォンの各々によって取得される受信信号に含まれる音声の発生源である音声源が近傍界に位置する場合には、前記受信信号を球面波として扱って前記音声源の方位を判定し、前記音声源が前記遠方界に位置する場合には、前記受信信号を平面波として扱って前記音声源の方位を判定する音声源方位判定部と、前記音声源の前記方位を含む方位範囲以外の方位範囲から到来する音を抑圧するようにビームフォーミングを行うビームフォーミング処理部とを有することを特徴とする音声処理装置が提供される。

【発明の効果】

【0008】

本発明によれば、音声源が近傍界に位置する場合には、音声を球面波として扱うため、音声源が近傍界に位置する場合であっても、音声源の方位を高精度に判定することができる。音声源の方位を高精度に判定し得るため、本発明によれば、目的音以外の音を確実に抑制することができる。しかも、音声源が遠方界に位置する場合には、音声を平面波として扱って音声源の方位を判定するため、音声源の方位を判定するための処理負荷を軽くすることができる。従って、本発明によれば、音声認識の確実性を向上し得る良好な音声処理装置を提供することができる。

【図面の簡単な説明】

【0009】

【図1】車両の構成を示す概略図である。

【図2】本発明の一実施形態による音声処理装置のシステム構成を示すブロック図である。

【図3】マイクロフォンの配置の例を示す概略図である。

【図4】音声源が遠方界に位置する場合と近傍界に位置する場合とを示す図である。

【図5】音楽の除去のアルゴリズムを示す概略図である。

【図6】音楽の除去前と除去後の信号波形を示す図である。

【図7】音声源の方位の判定のアルゴリズムを示す図である。

【図8】適応フィルタ係数、音声源の方位角、及び、音声信号の振幅を示す図である。

【図9】ビームフォーマの指向性を概念的に示す図である。

【図10】ビームフォーマのアルゴリズムを示す図である。

【図11】ビームフォーマにより得られる指向性の例を示すグラフである。

【図12】ビームフォーマと音声源方位判定キャンセル処理とを組み合わせた場合の角度特性を示す図である。

【図13】ビームフォーマにより得られる指向性の例を示すグラフである。

【図14】ノイズの除去のアルゴリズムを示す図である。

【図15】ノイズの除去前と除去後の信号波形を示す図である。

【図16】本発明の一実施形態による音声処理装置の動作を示すフローチャートである。

【発明を実施するための形態】

【0010】

以下、本発明の実施の形態について図面を用いて説明する。なお、本発明は以下の実施形態に限定されるものではなく、その要旨を逸脱しない範囲において適宜変更可能である。また、以下で説明する図面において、同じ機能を有するものは同一の符号を付し、その説明を省略又は簡潔にすることもある。

10

20

30

40

50

【 0 0 1 1 】

[一実施形態]

本発明の一実施形態による音声処理装置を図 1 乃至図 1 6 を用いて説明する。

【 0 0 1 2 】

本実施形態による音声処理装置について説明するに先立って、車両の構成について図 1 を用いて説明する。図 1 は、車両の構成を示す概略図である。

【 0 0 1 3 】

図 1 に示すように、車両（自動車）の車体（車室）4 6 の前部には、運転者用の座席である運転席 4 0 と助手席者用の座席である助手席 4 4 とが配されている。運転席 4 0 は、例えば車室 4 6 の右側に位置している。運転席 4 0 の前方には、ステアリングホイール（ハンドル）7 8 が配されている。助手席 4 4 は、例えば車室 4 6 の左側に位置している。運転席 4 0 と助手席 4 4 とにより、前部座席が構成されている。運転席 4 0 の近傍には、運転者が音声を発する場合における音声源 7 2 a が位置する。助手席 4 4 の近傍には、助手席者が音声を発する場合における音声源 7 2 b が位置する。運転者も助手席者も座席 4 0 , 4 4 に着座した状態で上半身を動かし得るため、音声源 7 2 の位置は変化し得る。車体 4 6 の後部には、後部座席 7 0 が配されている。なお、ここでは、個々の音声源を区別しないで説明する場合には、符号 7 2 を用い、個々の音声源を区別して説明する場合には、符号 7 2 a、7 2 b を用いることとする。

10

【 0 0 1 4 】

前部座席 4 0 , 4 4 の前方には、複数のマイクロフォン 2 2 (2 2 a ~ 2 2 c)、即ち、マイクロフォンアレイが配されている。なお、ここでは、個々のマイクロフォンを区別しないで説明する場合には、符号 2 2 を用い、個々のマイクロフォンを区別して説明する場合には、符号 2 2 a ~ 2 2 c を用いることとする。マイクロフォン 2 2 は、ダッシュボード 4 2 に配されていてもよいし、ルーフに近い部位に配されていてもよい。

20

【 0 0 1 5 】

前部座席 4 0 , 4 4 の音声源 7 2 とマイクロフォン 2 2 との間の距離は、数十 c m 程度である場合が多い。しかし、マイクロフォン 2 2 と音声源 7 2 との間の距離は、数十 c m より小さくなることもあり得る。また、マイクロフォン 2 2 と音声源 7 2 との間の距離は、1 m を超えることもあり得る。

【 0 0 1 6 】

車体 4 6 の内部には、車載音響機器（カーオーディオ機器）8 4（図 2 参照）のスピーカシステムを構成するスピーカ（ラウドスピーカ）7 6 が配されている。スピーカ 7 6 から発せられる音楽（ミュージック）は、音声認識を行う上でのノイズとなり得る。

30

【 0 0 1 7 】

車体 4 6 には、車両を駆動するためのエンジン 8 0 が配されている。エンジン 8 0 から発せられる音は、音声認識を行う上でのノイズとなり得る。

【 0 0 1 8 】

車両の走行中に路面の刺激によって車室 4 6 内に発生する騒音、即ち、ロードノイズも、音声認識を行う上でのノイズとなり得る。また、車両が走行する際に生ずる風切り音も、音声認識を行う上でのノイズ源となり得る。また、車体 4 6 の外部にも、ノイズ源 8 2 は存在し得る。外部ノイズ源 8 2 から発せられる音も、音声認識を行う上でのノイズとなり得る。

40

【 0 0 1 9 】

車体 4 6 に配された様々な機器に対する操作を、音声による指示によって行い得ると便利である。音声による指示は、例えば図示しない自動音声認識装置を用いて認識される。本実施形態による音声処理装置は、音声認識の精度の向上に資するものである。

【 0 0 2 0 】

図 2 は、本実施形態による音声処理装置のシステム構成を示すブロック図である。

【 0 0 2 1 】

図 2 に示すように、本実施形態による音声処理装置は、前処理部 1 0 と、処理部 1 2 と

50

、後処理部 14 と、音声源方位判定部 16 と、適応アルゴリズム決定部 18 と、ノイズモデル決定部 20 とを含む。

【0022】

本実施形態による音声処理装置が更に図示しない自動音声認識装置を含んでいてもよいし、本実施形態による音声処理装置と自動音声認識装置とが別個の装置であってもよい。これらの構成要素と自動音声認識装置とを含む装置は、音声処理装置と称することもできるし、自動音声認識装置と称することもできる。

【0023】

前処理部 10 には、複数のマイクロフォン 22a ~ 22c の各々によって取得される信号、即ち、受信信号が入力されるようになっている。マイクロフォン 22 としては、例えば、無指向性のマイクロフォンが用いられる。

10

【0024】

図 3 は、マイクロフォンの配置の例を示す概略図である。図 3 (a) は、マイクロフォン 22 の数が 3 個の場合を示している。図 3 (b) は、マイクロフォン 22 の数が 2 個の場合を示している。複数のマイクロフォン 22 は、直線上に位置するように配されている。

【0025】

図 4 は、音声源が遠方界に位置する場合と近傍界に位置する場合とを示す図である。図 4 (a) は、音声源 72 が遠方界に位置する場合を示しており、図 4 (b) は、音声源 72 が近傍界に位置する場合を示している。d は、音声源 72 からマイクロフォン 22 までの距離の差を示している。θ は、音声源 72 の方位を示している。

20

【0026】

図 4 (a) に示すように、音声源 72 が遠方界に位置する場合には、マイクロフォン 22 に到達する音声は、平面波とみなすことができる。このため、本実施形態では、音声源 72 が遠方界に位置する場合には、マイクロフォン 22 に到達する音声を平面波として取り扱って、音声源 72 の方位 (方向)、即ち、音源方位 (DOA : Direction Of Arrival) を判定する。マイクロフォン 22 に到達する音声を平面波として扱うことが可能なため、音声源 72 が遠方界に位置する場合には、2 個のマイクロフォン 22 を用いて音声源 72 の方位を判定し得る。なお、音声源 72 の位置やマイクロフォン 22 の配置によっては、マイクロフォン 22 の数が 2 個の場合であっても、近傍界に位置する音声源 72 の方位を判定し得る。

30

【0027】

図 4 (b) に示すように、音声源 72 が近傍界に位置する場合には、マイクロフォン 22 に到達する音声は、球面波とみなすことができる。このため、本実施形態では、音声源 72 が近傍界に位置する場合には、マイクロフォン 22 に到達する音声を球面波として扱うことを要するため、音声源 72 が近傍界に位置する場合には、少なくとも 3 個のマイクロフォン 22 を用いて音声源 72 の方位を判定する。ここでは、説明の簡略化のため、マイクロフォン 22 の数を 3 個とする場合を例に説明する。

【0028】

マイクロフォン 22a とマイクロフォン 22b との距離 L1 は、比較的長く設定されている。マイクロフォン 22b とマイクロフォン 22c との距離 L2 は、比較的短く設定されている。

40

【0029】

本実施形態において距離 L1 と距離 L2 とを異ならせているのは、以下のような理由によるものである。即ち、本実施形態では、各々のマイクロフォン 22 に到達する音声 (受信信号の到来時間差 (TDOA : Time Delay Of Arrival)) に基づいて、音声源 72 の方位を特定する。周波数が比較的低い音声は波長が比較的長いため、周波数が比較的低い音声に対応するためには、マイクロフォン 22 間の距離を比較的大きく設定することが好ましい。このため、本実施形態では、マイクロフォン 22a とマイクロフォン 22b との間

50

の距離 L_1 を比較的長く設定している。一方、周波数が比較的高い音声は波長が比較的短いため、周波数が比較的高い音声に対応するためには、マイクロフォン 22 間の距離を比較的小さく設定することが好ましい。そこで、本実施形態では、マイクロフォン 22 b とマイクロフォン 22 c との間の距離 L_2 を比較的短く設定している。

【0030】

マイクロフォン 22 a とマイクロフォン 22 b との間の距離 L_1 は、例えば 3400 Hz 以下の周波数の音声に対して好適とすべく、例えば 5 cm 程度とする。マイクロフォン 22 b とマイクロフォン 22 c との間の距離 L_2 は、例えば 3400 Hz を超える周波数の音声に対して好適とすべく、例えば 2.5 cm 程度とする。なお、距離 L_1 、 L_2 は、これらに限定されるものではなく、適宜設定し得る。

10

【0031】

本実施形態において、音声源 72 が遠方界に位置する場合に、マイクロフォン 22 に到達する音声を平面波として扱うのは、音声を平面波として扱う場合の方が、音声を球面波として扱う場合よりも、音声源 72 の方位を判定するための処理が簡略なためである。このため、本実施形態では、音声源 72 が遠方界に位置する場合には、マイクロフォン 22 に到達する音声を平面波として扱う。マイクロフォン 22 に到達する音声を平面波として扱うため、遠方界に位置する音声源 72 の方位を判定する際には、音声源 72 の方位を判定するための処理の負荷を軽くすることができる。

【0032】

なお、音声源 72 の方位を判定するための処理の付加は重くなるが、音声源 72 が近傍界に位置する場合には、マイクロフォン 22 に到達する音声を球面波として扱う。音声源 72 が近傍界に位置する場合には、マイクロフォン 22 に到達する音声を球面波として扱わないと、音声源 72 の方位を正確に判定し得ないためである。

20

【0033】

このように、本実施形態では、音声源 72 が遠方界に位置する場合には、音声を平面波として扱って音声源 72 の方位を判定し、音声源 72 が近傍界に位置する場合には、音声を球面波として扱って音声源 72 の方位を判定する。

【0034】

図 2 に示すように、複数のマイクロフォン 22 によって取得される受信信号が、前処理部 10 に入力されるようになっている。前処理部 10 では、音場補正が行われる。音場補正においては、音響空間である車室 46 の音響特性を考慮したチューニングが行われる。

30

【0035】

マイクロフォン 22 によって取得される受信信号に音楽が含まれている場合には、前処理部 10 は、マイクロフォン 22 によって取得される受信信号から音楽を除去する。前処理部 10 には、参照用音楽信号（参照信号）が入力されるようになっている。前処理部 10 は、マイクロフォン 22 によって取得される受信信号に含まれている音楽を、参照用音楽信号を用いて除去する。

【0036】

図 5 は、音楽の除去のアルゴリズムを示す概略図である。車載音響機器 84 によって音楽が再生されている際には、マイクロフォン 22 によって取得される受信信号には音楽が含まれる。マイクロフォン 22 によって取得される音楽を含む受信信号は、前処理部 10 内に設けられた音楽除去処理部 24 に入力されるようになっている。また、参照用音楽信号が、音楽除去処理部 24 に入力されるようになっている。参照用音楽信号は、例えば、車載音響機器 84 のスピーカ 76 から出力された音楽を、マイクロフォン 26 a、26 b によって取得することにより得ることが可能である。また、スピーカ 76 によって音に変換される前の音楽ソース信号を、参照用音楽信号として、音楽除去処理部 24 に入力するようにしてもよい。

40

【0037】

音楽除去処理部 24 からの出力信号は、前処理部 10 内に設けられたステップサイズ判定部 28 に入力されるようになっている。ステップサイズ判定部 28 は、音楽除去処理部

50

24の出力信号のステップサイズの判定を行うものである。ステップサイズ判定部28によって判定されたステップサイズは、音楽除去処理部24にフィードバックされるようになっている。音楽除去処理部24は、参照用音楽信号を用い、ステップサイズ判定部28により判定されたステップサイズに基づき、周波数領域の正規化最小二乗法(NLMS: Normalized Least-Mean Square)のアルゴリズムによって、音楽を含む信号から音楽を除去する。車室46内における音楽の反響成分をも十分に除去すべく、十分な処理段数で音楽の除去の処理が行われる。

【0038】

図6は、音楽の除去前と除去後の信号波形を示す図である。横軸は時間を示しており、縦軸は振幅を示している。グレーで示した信号は音楽の除去前を示しており、ブラックで示した信号は音楽の除去後を示している。図6から分かるように、音楽が確実に除去されている。

10

【0039】

このようにして音楽が除去された信号が、前処理部10の音楽除去処理部24から出力され、処理部12に入力される。なお、前処理部10において音楽を十分に除去し得ない場合には、後処理部14においても、音楽の除去の処理を行うようにしてもよい。

【0040】

音声源方位判定部16では、音声源の方位の判定が行われる。図7は、音声源の方位の判定のアルゴリズムを示す図である。複数のマイクロフォン22のうちのあるマイクロフォン22からの信号が、音声源方位判定部16内に設けられた遅延部30に入力されるようになっている。複数のマイクロフォン22のうち他のマイクロフォン22からの信号が、音声源方位判定部16内に設けられた適応フィルタ32に入力されるようになっている。遅延部30の出力信号と適応フィルタ32の出力信号とが、減算点34に入力されるようになっている。減算点34においては、遅延部30の出力信号から適応フィルタ34の出力信号が減算される。減算点34において減算処理が行われた信号に基づいて、適応フィルタ32が調整される。適応フィルタ32からの出力は、ピーク検出部36に入力されるようになっている。ピーク検出部36は、適応フィルタ係数のピーク(最大値)を検出するものである。適応フィルタ係数のピークに対応する到来時間差が、目的音の到来方位に対応する到来時間差である。従って、こうして求められた到来時間差に基づいて、音声源72の方位、即ち、目的音の到来方位を判定することが可能となる。

20

30

【0041】

音の速度を c [m/s]、マイクロフォン間の距離を d [m]、到来時間差を [秒] とすると、音声源72の方向 [度] は、以下のような式(1)によって表される。なお、音速 c は、340 [m/s] 程度である。

【0042】

$$= (180 /) \times \arccos (\cdot c / d) \cdot \cdot \cdot (1)$$

【0043】

図8は、適応フィルタ係数、音声源の方位角、及び、音声信号の振幅を示す図である。図8(a)では、適応フィルタ係数がピークとなる部分にハッチングを付している。図8(b)は、到来時間差に基づいて判定された音声源72の方位を示している。図8(c)は、音声信号の振幅を示している。なお、図8は、運転者と助手席者とで交互に音声を発した場合を示している。ここでは、運転者が音声を発する場合の音声源72aの方位は、1とした。助手席者が音声を発する場合の音声源72bの方位は、2とした。

40

【0044】

図8(a)に示すように、適応フィルタ係数 $w(t,)$ のピークに基づいて、到来時間差を検出することが可能である。運転者が音声を発した場合には、適応フィルタ係数のピークに対応する到来時間差は、例えば $-t_1$ 程度となる。そして、到来時間差に基づいて音声源72aの方位角を判定すると、音声源72aの方位角は例えば1程度と判定される。一方、助手席者が音声を発した場合には、適応フィルタ係数のピークに対応する到来時間差は、例えば t_2 程度となる。そして、到来時間差に基づいて音声源7

50

2 b の方位角を判定すると、音声源 7 2 b の方位角は例えば 2 度程度と判定される。なお、ここでは、1 の方位に運転者が位置しており、2 の方位に助手席者が位置している場合を例に説明したが、これに限定されるものではない。音声源 7 2 が近傍界に位置する場合であっても、音声源 7 2 が遠方界に位置する場合であっても、到来時間差に基づいて、音声源 7 2 の位置を特定することが可能である。但し、音声源 7 2 が近傍界に位置する場合には、上述したように、マイクロフォン 2 2 が 3 個以上必要であるため、音声源 7 2 の方位を求めるための処理の負荷は重くなる。

【0045】

音声源方位判定部 1 6 の出力信号、即ち、音声源 7 2 の方位を示す信号が、適応アルゴリズム決定部 1 8 に入力されるようになってきている。適応アルゴリズム決定部 1 8 は、音声源 7 2 の方位に基づいて適応アルゴリズムを決定するものである。適応アルゴリズム決定部 1 8 によって決定された適応アルゴリズムを示す信号が、適応アルゴリズム決定部 1 8 から処理部 1 2 に入力されるようになってきている。

10

【0046】

処理部 1 2 は、適応的に指向性を形成する信号処理である適応ビームフォーミングを行うものである（適応ビームフォーマ）。ビームフォーマとしては、例えば Frost ビームフォーマを用いることができる。なお、ビームフォーミングは、Frost ビームフォーマに限定されるものではなく、様々なビームフォーマを適宜適用することができる。処理部 1 2 は、適応アルゴリズム決定部 1 8 によって決定された適応アルゴリズムに基づいて、ビームフォーミングを行う。本実施形態において、ビームフォーミングを行うのは、目的音の到来方位に対しての感度を確保しつつ、目的音の到来方向以外の感度を低下させるためである。目的音は、例えば運転者から発せられる音声である。運転者は運転席 4 0 に着座した状態で上半身を動かす得るため、音声源 7 2 a の位置は変化し得る。音声源 7 2 a の位置の変化に応じて、目的音の到来方位は変化する。良好な音声認識を行うためには、目的音の到来方向以外の感度を確実に低下させることが好ましい。そこで、本実施形態では、上記のようにして判定される音声源 7 2 の方位に基づいて、当該方位を含む方位範囲以外の方位範囲からの音声を抑圧すべく、ビームフォーマを順次更新する。

20

【0047】

図 9 は、ビームフォーマの指向性を概念的に示す図である。図 9 は、音声認識の対象とすべき音声源 7 2 a が運転席 4 0 に位置している場合のビームフォーマの指向性を概念的に示している。図 9 におけるハッチングは、到来音が抑圧（抑制、低減）される方位範囲を示している。図 9 に示すように、運転席 4 0 の方位を含む方位範囲以外の方位範囲から到来する音が抑圧される。

30

【0048】

なお、音声認識の対象とすべき音声源 7 2 b が助手席 4 4 に位置している場合には、助手席 4 4 の方位を含む方位範囲以外の方位範囲から到来する音が抑圧されるようにすればよい。

【0049】

図 10 は、ビームフォーマのアルゴリズムを示す図である。マイクロフォン 2 2 a ~ 2 2 c によって取得される受信信号が、前処理部 1 0（図 2 参照）を介して、処理部 1 2 内に設けられた窓関数 / 高速フーリエ変換処理部 4 8 a ~ 4 8 c にそれぞれ入力されるようになってきている。窓関数 / 高速フーリエ変換処理部 4 8 a ~ 4 8 c は、窓関数処理及び高速フーリエ変換処理を行うものである。本実施形態において、窓関数処理及び高速フーリエ変換処理を行うのは、周波数領域での計算は時間領域での計算より速いためである。窓関数 / 高速フーリエ変換処理部 4 8 a の出力信号 $X_{1,k}$ とビームフォーマの重みテンソル $W_{1,k}^*$ とが、乗算点 5 0 a において乗算されるようになってきている。窓関数 / 高速フーリエ変換処理部 4 8 b の出力信号 $X_{2,k}$ とビームフォーマの重みテンソル $W_{2,k}^*$ とが、乗算点 5 0 b において乗算されるようになってきている。窓関数 / 高速フーリエ変換処理部 4 8 c の出力信号 $X_{3,k}$ とビームフォーマの重みテンソル $W_{3,k}^*$ とが、乗算点 5 0 c において乗算されるようになってきている。乗算点 5 0 a ~ 5 0 c においてそれぞれ乗算

40

50

処理された信号が、加算点 5 2 において加算されるようになっている。加算点 5 2 において加算処理された信号 Y_k は、処理部 1 2 内に設けられた逆高速フーリエ変換 / 重畳加算処理部 5 4 に入力されるようになっている。逆高速フーリエ変換 / 重畳加算処理部 5 4 は、逆高速フーリエ変換処理及び重畳加算 (OLA: OverLap-Add) 法による処理を行うものである。重畳加算法による処理を行うことにより、周波数領域の信号が時間領域の信号に戻される。逆高速フーリエ変換処理及び重畳加算法による処理が行われた信号が、逆高速フーリエ変換 / 重畳加算処理部 5 4 から後処理部 1 4 に入力されるようになっている。

【0050】

図 1 1 は、ビームフォーマにより得られた指向性 (角度特性) を示す図である。横軸は方位角を示しており、縦軸は出力信号パワーを示している。図 1 1 から分かるように、例えば方位角 1 と方位角 2 とにおいて出力信号パワーが極小となる。方位角 1 と方位角 2 との間においても、十分な抑圧が行われている。図 1 1 に示すような指向性のビームフォーマを用いれば、助手席から到来する音を十分に抑圧することができる。一方、運転席から到来する音は、殆ど抑圧されることなくマイクロフォン 2 2 に到達する。

10

【0051】

本実施形態では、音声源 7 2 から到来する音声の大きさよりも、音声源 7 2 の方位を含む方位範囲以外の方位範囲から到来する音の方が大きい場合には、音声源 7 2 の方位の判定を中断する (音声源方位判定キャンセル処理)。例えば、運転者からの音声を取得するようにビームフォーマが設定されている場合において、運転者からの音声よりも助手席者からの音声が大きい場合には、音声源の方位の推定を中断する。この場合、マイクロフォン 2 2 によって取得される受信信号を十分に抑圧する。図 1 2 は、ビームフォーマと音声源方位判定キャンセル処理とを組み合わせた場合の指向性 (角度特性) を示す図である。実線は、ビームフォーマの指向性を示している。一点鎖線は、音声源方位判定キャンセル処理の角度特性を示している。例えば 1 より小さい方位から到来する音声、又は、例えば 2 より大きい方位から到来する音声が、運転者からの音声よりも大きい場合には、音声源方位判定キャンセル処理が行われる。なお、ここでは、運転者からの音声を取得するようにビームフォーマが設定されている場合を例に説明したが、助手席者からの音声を取得するようにビームフォーマが設定されていてもよい。この場合には、助手席者からの音声よりも運転者からの音声が大きい場合には、音声源の方位の推定を中断する。

20

【0052】

図 1 3 は、マイクロフォンが 2 個の場合におけるビームフォーマにより得られる指向性を示すグラフである。横軸は方位角であり、縦軸は出力信号パワーである。マイクロフォン 2 2 が 2 個であるため、極小値となる角度が 1 箇所のみである。図 1 3 から分かるように、例えば方位角 1 においては著しい抑圧が可能であるが、音声源 7 2 の方位の変化に対するロバスト性はあまり高くない。

30

【0053】

こうして、音声源 7 2 の方位を含む方位範囲以外の方位範囲から到来する音が抑圧された信号が、処理部 1 2 から出力される。処理部 1 2 からの出力信号は、後処理部 1 4 に入力されるようになっている。

【0054】

後処理部 (後処理適応フィルタ) 1 4 においては、ノイズの除去が行われる。かかるノイズとしては、例えばエンジンノイズ、ロードノイズ、風切り音等が挙げられる。図 1 4 は、ノイズの除去のアルゴリズムを示す図である。ノイズモデル決定部 2 0 内に設けられた基本波判定部 5 6 によって、ノイズの基本波が判定される。基本波判定部 5 6 は、ノイズの基本波に基づいた正弦波を出力する。基本波判定部 5 6 から出力される正弦波は、ノイズモデル決定部 2 0 内に設けられたモデリング処理部 5 8 に入力されるようになっている。モデリング処理部 5 8 は、非線形マッピング処理部 6 0 と、線形フィルタ 6 2 と、非線形マッピング処理部 6 4 とを有している。モデリング処理部 5 8 は、Hammerstein-Wien er 非線形モデルによるモデリング処理を行うものである。モデリング処理部 5 8 には、非線形マッピング処理部 6 0、線形フィルタ 6 2 及び非線形マッピング処理部 6 4 が設けら

40

50

れている。モデリング処理部 5 8 は、基本波判定部 5 6 から出力される正弦波に対してモデリング処理を行うことにより、参照用ノイズ信号を生成する。モデリング処理部 5 8 から出力される参照用ノイズ信号は、ノイズが含まれた信号からノイズを除去するための参照信号となる。参照用ノイズ信号は、後処理部 1 4 内に設けられたノイズ除去処理部 6 6 に入力されるようになっていて、ノイズ除去処理部 6 6 には、処理部 1 2 からのノイズを含む信号も入力されるようになっていて、ノイズ除去処理部 6 6 は、参照用ノイズ信号を用い、正規化最小二乗法のアルゴリズムによって、ノイズを含む信号からノイズを除去する。ノイズ除去処理部 6 6 からは、ノイズが除去された信号が出力される。

【 0 0 5 5 】

図 1 5 は、ノイズの除去前と除去後の信号波形を示す図である。横軸は時間を示しており、縦軸は振幅を示している。グレーで示した信号はノイズ除去前を示しており、ブラックで示した信号はノイズ除去後を示している。図 1 5 から分かるように、ノイズが確実に除去されている。

10

【 0 0 5 6 】

後処理部 1 4 においては、歪低減処理も行われる。なお、ノイズの除去は、後処理部 1 4 においてのみ行われるわけではない。マイクロフォン 2 2 を介して取得された音に対して、前処理部 1 0、処理部 1 2 及び後処理部 1 4 において行われる一連の処理によって、ノイズの除去が行われる。

【 0 0 5 7 】

こうして、後処理部 1 4 によって後処理が行われた信号が、図示しない自動音声認識装置に音声出力として出力される。目的音以外の音が抑圧された良好な目的音が自動音声認識装置に入力されるため、自動音声認識装置は、音声認識の精度を向上することができる。自動音声認識装置による音声認識結果に基づいて、車両に搭載されている機器等に対しての操作が自動で行われる。

20

【 0 0 5 8 】

次に、本実施形態による音声処理装置の動作について図 1 6 を用いて説明する。図 1 6 は、本実施形態による音声処理装置の動作を示すフローチャートである。

【 0 0 5 9 】

まず、音声処理装置の電源が ON にされる (ステップ S 1)。

【 0 0 6 0 】

次に、乗員による呼びかけが音声処理装置に対して行われる (ステップ S 2)。かかる呼びかけによって、音声処理が開始される。ここでは、例えば、運転者によって呼びかけが行われる場合を例に説明する。なお、呼びかけは、運転者が行わなくてもよい。例えば、助手席者が呼びかけを行ってもよい。また、呼びかけは、特定の言葉であってもよいし、単なる発声であってもよい。

30

【 0 0 6 1 】

次に、呼びかけを行った音声源 7 2 の方位が判定される (ステップ S 3)。音声源 7 2 の方位の判定は、上述したように、音声源方位判定部 1 6 等によって行われる。

【 0 0 6 2 】

次に、音声源 7 2 の方位に応じて、ビームフォーマの指向性を設定する (ステップ S 4)。
)。ビームフォーマの指向性の設定は、上述したように、適応アルゴリズム決定部 1 8、
処理部 1 2 等によって行われる。

40

【 0 0 6 3 】

音声源 7 2 の方位を含む所定の方位範囲以外の方位範囲から到来する音の大きさが、音声源 7 2 から到来する音声の大きさ以上である場合には (ステップ S 5 において YES)、音声源 7 2 の判定を中断する (ステップ S 6)。

【 0 0 6 4 】

一方、音声源 7 2 の方位を含む所定の方位範囲以外の方位範囲から到来する音の大きさが、音声源 7 2 から到来する音声の大きさ以上でない場合には (ステップ S 5 において NO)、ステップ S 3、S 4 を繰り返し行う。

50

【 0 0 6 5 】

こうして、音声源 7 2 の位置の変化に応じて、ビームフォーマが適応的に設定され、目的音以外の音が確実に抑制される。

【 0 0 6 6 】

このように、本実施形態によれば、音声源 7 2 が近傍界に位置する場合には、音声を球面波として扱うため、音声源 7 2 が近傍界に位置する場合であっても、音声源 7 2 の方位を高精度に判定することができる。音声源 7 2 の方位を高精度に判定し得るため、本実施形態によれば、目的音以外の音を確実に抑制することができる。しかも、音声源 7 2 が遠方界に位置する場合には、音声を平面波として扱って音声源 7 2 の方位を判定するため、音声源 7 2 の方位を判定するための処理負荷を軽くすることができる。従って、本実施形態によれば、音声認識の確実性を向上し得る良好な音声処理装置を提供することができる。

10

【 0 0 6 7 】

また、本実施形態によれば、受信信号に含まれる音楽を除去する音楽除去処理部 2 4 が設けられているため、車載音響機器 8 4 から音楽が再生されている場合であっても、良好な音声認識を行うことが可能となる。

【 0 0 6 8 】

また、本実施形態によれば、受信信号に含まれるノイズを除去するノイズ除去処理部 6 6 が設けられているため、車両が走行中であっても、良好な音声認識を行うことが可能となる。

20

【 0 0 6 9 】

[変形実施形態]

上記実施形態に限らず種々の変形が可能である。

【 0 0 7 0 】

例えば、上記実施形態では、マイクロフォン 2 2 の数が 3 個である場合を例に説明したが、マイクロフォン 2 2 の数は 3 個に限定されるものではなく、4 個以上であってもよい。多くのマイクロフォン 2 2 を用いれば、音声源 7 2 の方位をより高精度に判定し得る。

【 0 0 7 1 】

また、上記実施形態では、本実施形態による音声処理装置の出力が自動音声認識装置に入力される場合、即ち、本実施形態による音声処理装置の出力が音声認識に用いられる場合を例に説明したが、これに限定されるものではない。本実施形態による音声処理装置の出力が、自動音声認識に用いられなくてもよい。例えば、本実施形態による音声処理装置を、電話での会話における音声処理に適用してもよい。具体的には、本実施形態による音声処理装置を用いて目的音以外の音を抑圧し、良好な音声を送信するようにしてもよい。本実施形態による音声処理装置を電話での会話に適用すれば、良好な音声での通話を実現することができる。

30

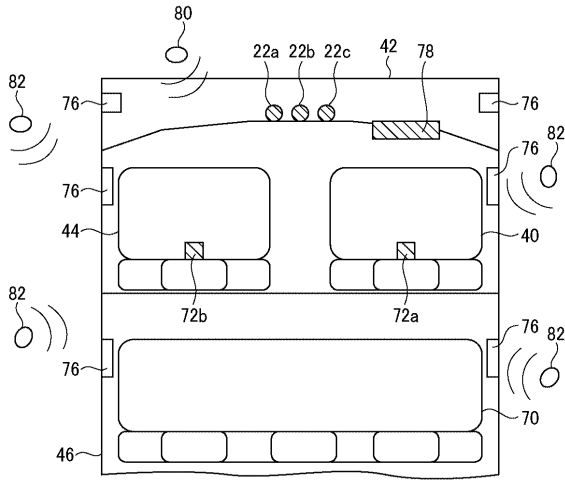
【 符号の説明 】

【 0 0 7 2 】

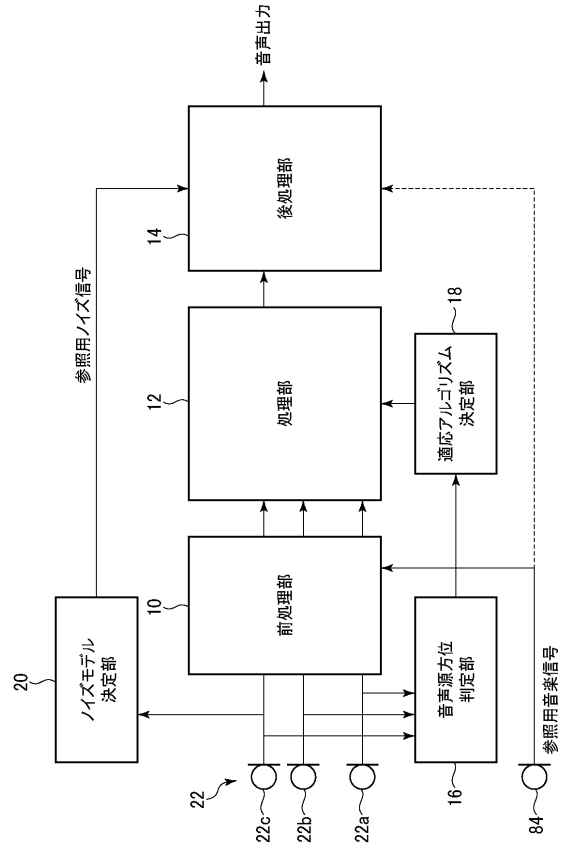
2 2 , 2 2 a ~ 2 2 c ... マイクロフォン
 4 0 ... 運転席
 4 2 ... ダッシュボード
 4 4 ... 助手席
 4 6 ... 車体
 7 2 , 7 2 a , 7 2 b ... 音声源
 7 6 ... スピーカ
 7 8 ... ステアリングホイール
 8 0 ... エンジン
 8 2 ... 外部ノイズ源
 8 4 ... 車載音響機器

40

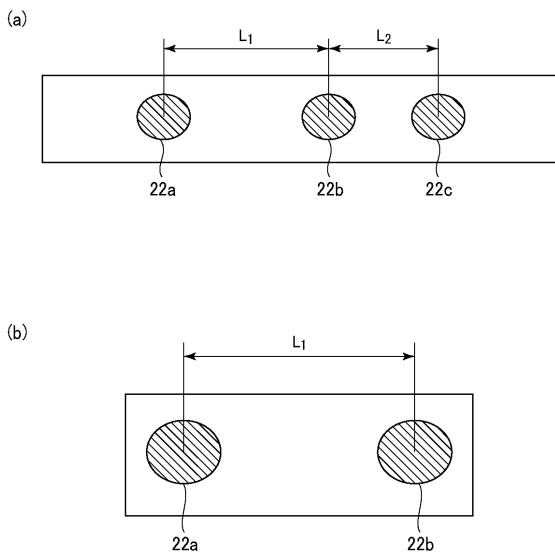
【 図 1 】



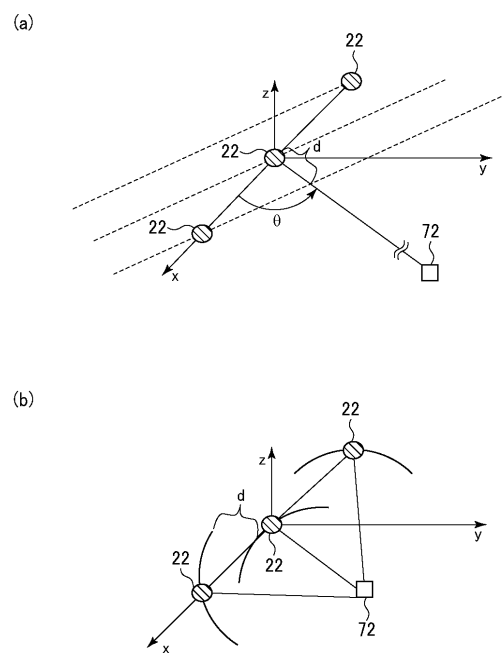
【 図 2 】



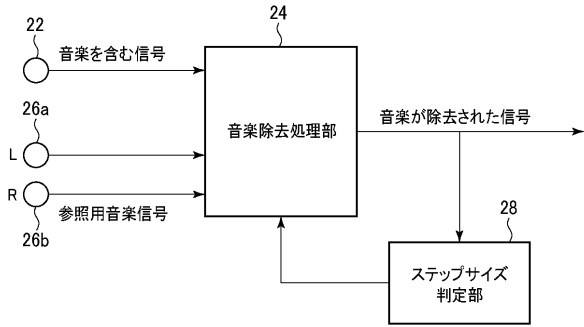
【 図 3 】



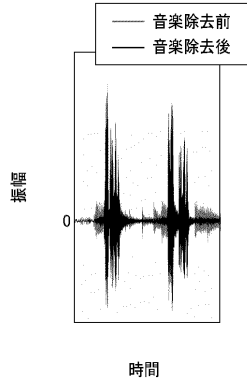
【 図 4 】



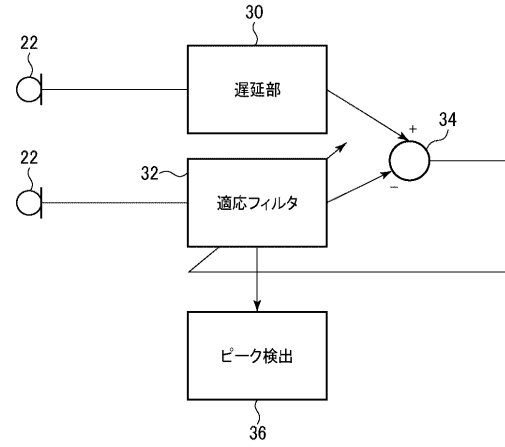
【図5】



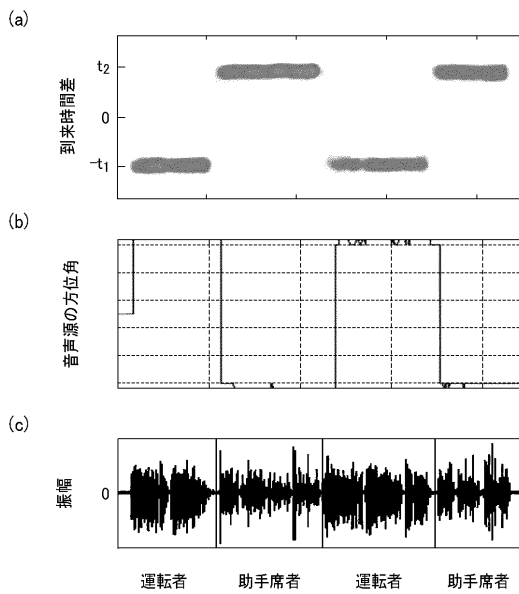
【図6】



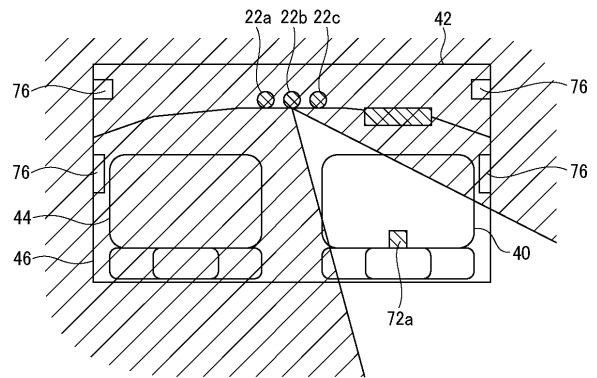
【図7】



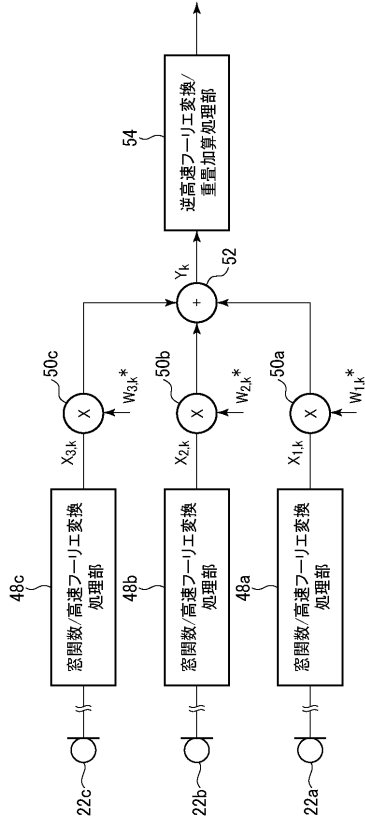
【図8】



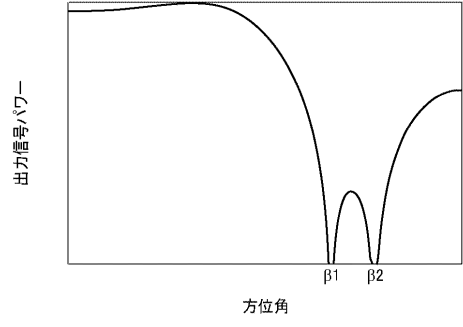
【図9】



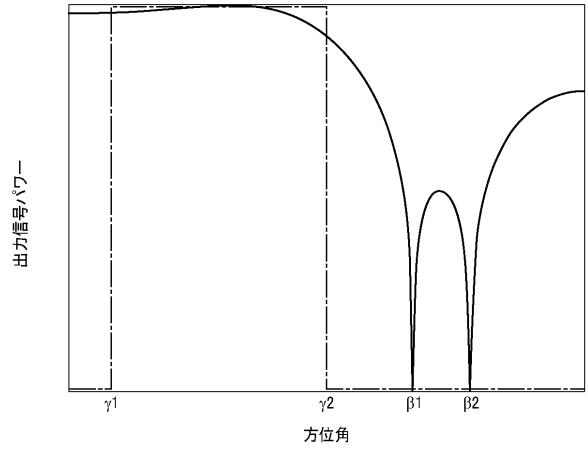
【図10】



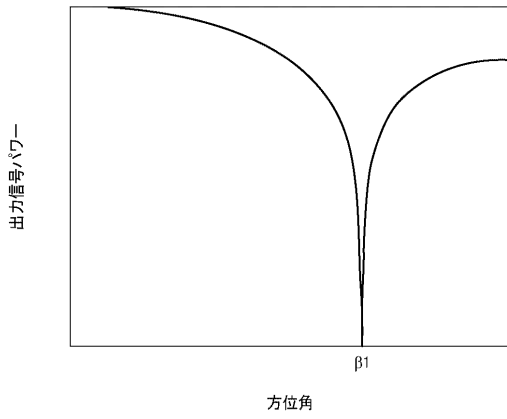
【図11】



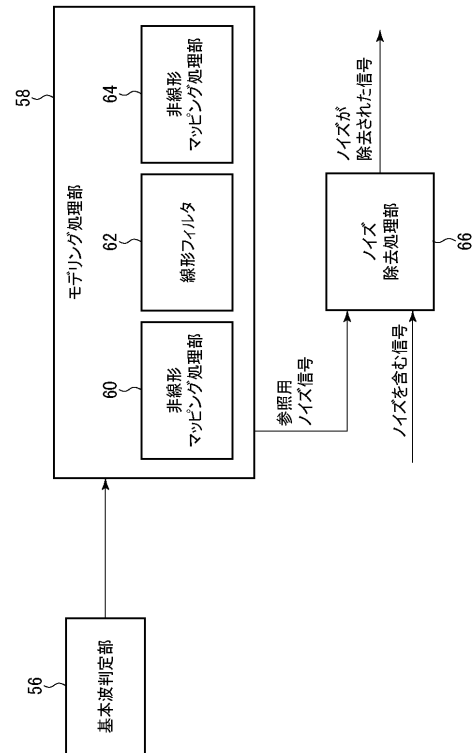
【図12】



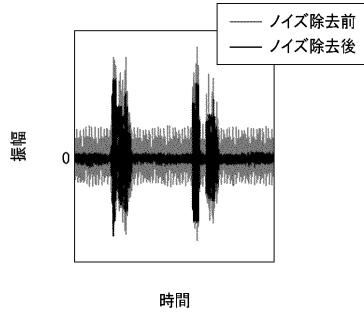
【図13】



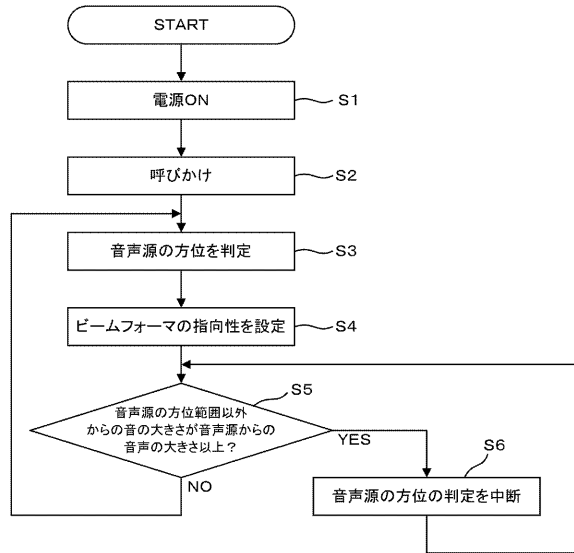
【図14】



【図 15】



【図 16】



フロントページの続き

(51)Int.Cl.

F I

テーマコード(参考)

G 1 0 L 15/00 2 0 0 J