

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第4554044号
(P4554044)

(45) 発行日 平成22年9月29日 (2010. 9. 29)

(24) 登録日 平成22年7月23日 (2010. 7. 23)

(51) Int. Cl.	F I
G 1 0 L 15/00 (2006. 01)	G 1 0 L 15/00 2 0 0 G
G 1 0 L 15/20 (2006. 01)	G 1 0 L 15/20 3 7 0 Z
G 1 0 L 15/28 (2006. 01)	G 1 0 L 15/28 2 3 0 K

請求項の数 14 (全 43 頁)

(21) 出願番号	特願2000-227787 (P2000-227787)	(73) 特許権者	000005821
(22) 出願日	平成12年7月27日 (2000. 7. 27)		パナソニック株式会社
(65) 公開番号	特開2001-100785 (P2001-100785A)		大阪府門真市大字門真1006番地
(43) 公開日	平成13年4月13日 (2001. 4. 13)	(74) 代理人	100098291
審査請求日	平成19年5月25日 (2007. 5. 25)		弁理士 小笠原 史朗
(31) 優先権主張番号	特願平11-213276	(72) 発明者	茨木 悟
(32) 優先日	平成11年7月28日 (1999. 7. 28)		大阪府門真市大字門真1006番地 松下
(33) 優先権主張国	日本国 (JP)		電器産業株式会社内
		(72) 発明者	金森 丈郎
			大阪府門真市大字門真1006番地 松下
			電器産業株式会社内
		(72) 発明者	河村 岳
			大阪府門真市大字門真1006番地 松下
			電器産業株式会社内

最終頁に続く

(54) 【発明の名称】 A V機器用音声認識装置

(57) 【特許請求の範囲】

【請求項 1】

複数のスピーカを通じてマルチチャンネル音響を出力する A V 機器に用いられ、マイクロホンを通じて入力されるユーザ音声を認識して、当該 A V 機器に所定の処理動作を行わせるための音声認識装置であって、

前記複数のスピーカへ向かうマルチチャンネル信号をモノラル化するモノラル化手段と

、
前記マイクロホンの出力（以下、マイクロホン出力）と、前記モノラル化手段の出力（以下、モノラル化信号）とが与えられ、当該モノラル化信号に基づいて前記マルチチャンネル音響の反響音を推定して、当該マイクロホン出力から当該反響音を除去する 1 つのエコーキャンセラと、

前記 1 つのエコーキャンセラの出力（以下、エコーキャンセラ出力）に基づいて前記ユーザ音声を認識する音声認識手段と、

前記マルチチャンネル信号および前記モノラル化信号のいずれかを前記複数のスピーカへと入力するための切り替え手段と、

音声認識動作の起動を命令する起動命令手段と、

音声認識動作の終了を命令する終了命令手段と、

前記起動命令手段および前記終了命令手段からの命令に応じて、前記音声認識手段を動作状態および待機状態のいずれかに設定する状態設定手段とを備え、

前記切り替え手段は、

10

20

前記状態設定手段によって前記音声認識手段が待機状態に設定されている時、前記マルチチャンネル信号を前記複数のスピーカへと入力し、

前記状態設定手段によって前記音声認識手段が動作状態に設定されている時、前記モノラル化信号を前記複数のスピーカへと入力することを特徴とする、A V 機器用音声認識装置。

【請求項 2】

前記マルチチャンネル信号のモノラル度を判定するモノラル度判定手段と、

前記マルチチャンネル信号を任意のモノラル度にモノラル化する任意度モノラル化手段とをさらに備え、

前記モノラル化手段は、前記マルチチャンネル信号を完全にモノラル化し、

前記任意度モノラル化手段は、前記モノラル度判定手段の判定結果が予め定められたモノラル度よりも低い場合、前記マルチチャンネル信号を当該予め定められたモノラル度にモノラル化することを特徴とする、請求項 1 に記載の A V 機器用音声認識装置。

【請求項 3】

前記マルチチャンネル信号は、3 チャンネル以上の信号であり、

前記マルチチャンネル信号を 2 チャンネル化する 2 チャンネル化手段をさらに備え、

前記モノラル化手段は、前記 2 チャンネル化手段の出力（以下、2 チャンネル化信号）をモノラル化し、

前記切り替え手段は、前記マルチチャンネル信号、前記 2 チャンネル化信号および前記モノラル化信号のいずれかを前記複数のスピーカへと入力することを特徴とする、請求項 1 に記載の A V 機器用音声認識装置。

【請求項 4】

前記モノラル化信号と前記エコーキャンセラ出力とに基づいて前記ユーザ音声を検出する音声検出手段をさらに備え、

前記切り替え手段は、

前記状態設定手段が前記音声認識手段を待機状態に設定している時には、前記マルチチャンネル信号を前記複数のスピーカへと入力し、

前記状態設定手段が前記音声認識手段を動作状態に設定しているが、前記音声検出手段によって前記ユーザ音声を検出されていない時には、前記 2 チャンネル化信号を前記複数のスピーカへと入力し、

前記音声検出手段によって前記ユーザ音声を検出されている時には、前記モノラル化信号を前記複数のスピーカへと入力することを特徴とする、請求項 3 に記載の A V 機器用音声認識装置。

【請求項 5】

前記モノラル化信号と前記エコーキャンセラ出力とに基づいて、前記エコーキャンセラにおいて前記反響音が十分に打ち消されているかを監視する打ち消し監視手段と、

前記モノラル化信号と前記エコーキャンセラ出力とに基づいて前記ユーザ音声を検出する音声検出手段と、

前記マルチチャンネル信号を減衰させる減衰手段とをさらに備え、

前記減衰手段は、前記打ち消し監視手段の監視結果が打ち消し不十分を示している状態において、前記音声検出手段がユーザ音声を検出した時、前記マルチチャンネル信号を減衰させることを特徴とする、請求項 1 に記載の A V 機器用音声認識装置。

【請求項 6】

前記エコーキャンセラは、

前記複数のスピーカと前記マイクロホンとの間の反響路のインパルスレスポンスを推定して、当該推定インパルスレスポンスと、前記モノラル化信号とから前記反響音を算出する適応デジタルフィルタと、

前記マイクロホン出力から、前記適応デジタルフィルタの出力を減算する減算手段とを含む、請求項 1 に記載の A V 機器用音声認識装置。

【請求項 7】

前記切り替え手段によって前記複数のスピーカへの入力の前記マルチチャンネル信号から前記モノラル化信号へと切り替えられる際に、前記適応デジタルフィルタの適応を促進するためのモノラル適応音を発生する適応音発生手段をさらに備える、請求項6に記載のAV機器用音声認識装置。

【請求項 8】

前記適応デジタルフィルタの適応速度を制御する適応制御手段をさらに備え、

前記適応制御手段は、モノラル用の速い適応速度と、マルチチャンネル用の遅い適応速度とを持ち、前記状態設定手段が前記音声認識手段を動作状態に設定している時には速い適応速度を選択し、待機状態に設定している時には遅い適応速度を選択することを特徴とする、請求項6に記載のAV機器用音声認識装置。

10

【請求項 9】

前記適応制御手段へは、前記複数のスピーカに入力される信号がマルチチャンネル信号かモノラル信号かを示す識別信号が与えられ、

前記適応制御手段は、前記識別信号がモノラルを示す場合、前記状態設定手段が前記音声認識手段を動作状態に設定しているか待機状態に設定しているかに関わらず、前記速い適応速度を選択することを特徴とする、請求項8に記載のAV機器用音声認識装置。

【請求項 10】

前記マルチチャンネル信号のモノラル度を判定するモノラル度判定手段と、

前記モノラル度判定手段の判定結果に基づいて、前記適応デジタルフィルタの適応速度を制御する適応制御手段とをさらに備える、請求項6に記載のAV機器用音声認識装置。

20

【請求項 11】

前記適応制御手段は、前記マルチチャンネル信号のモノラル度が高いほど、前記適応デジタルフィルタの適応速度を速くすることを特徴とする、請求項10に記載のAV機器用音声認識装置。

【請求項 12】

不揮発メモリをさらに備え、

前記不揮発メモリは、

電源が"OFF"される時、前記適応デジタルフィルタによって推定されたインパルスレスポンスを取得して記憶し、

30

電源が"ON"される時、記憶している電源"OFF"時の推定インパルスレスポンスを前記適応デジタルフィルタに与え、

前記適応デジタルフィルタは、前記不揮発メモリから与えられた電源"OFF"時の推定インパルスレスポンスを初期値として、インパルスレスポンスの推定を開始することを特徴とする、請求項6に記載のAV機器用音声認識装置。

【請求項 13】

前記モノラル化信号と前記エコーキャンセラ出力とに基づいて前記ユーザ音声を検出する音声検出手段をさらに備え、

前記起動命令手段は、ボタンが押されると、前記状態設定手段へ起動命令を発するようなボタンスイッチであり、

40

前記終了命令手段は、前記音声検出手段がユーザ音声を検出しない状態が予め定められた時間以上継続すると、前記状態設定手段へ終了命令を発する時限スイッチであることを特徴とする、請求項1に記載のAV機器用音声認識装置。

【請求項 14】

前記モノラル化信号と前記エコーキャンセラ出力とに基づいて前記ユーザ音声を検出する音声検出手段をさらに備え、

前記起動命令手段は、前記音声検出手段がユーザ音声を検出すると、前記状態設定手段へ起動命令を発する音声スイッチであり、

前記終了命令手段は、前記音声検出手段がユーザ音声を検出しない状態が予め定められた時間以上継続すると、前記状態設定手段へ終了命令を発する時限スイッチであることを

50

特徴とする、請求項 1 に記載の A V 機器用音声認識装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、音声認識装置に関し、より特定的には、2チャンネルステレオを含むマルチチャンネルオーディオを再生するTV、ラジオ、オーディオシステムなどのAV機器に用いられ、スピーカからオーディオが拡声された状態であっても、音声によるAV機器の制御や、音声によるAV機器への情報入力等が可能であるAV機器用音声認識装置に関する。

【従来の技術】

従来、スピーカからオーディオが拡声された状態において音声認識を行う技術を記載した文献としては、特開平5-22779号公報（発明の名称「音声認識遠隔制御装置」）が知られている。

【0002】

図23は、上記公報に開示されている技術を用いた従来のAV機器用音声認識装置の構成を示すブロック図である。図23の音声認識装置は、1つのスピーカ201を持ったAV機器に用いられる。図23において、従来のAV機器用音声認識装置は、マイクロホン202と、音声認識部203と、エコーキャンセラ204とを備えている。

【0003】

以上のように構成された従来のAV機器用音声認識装置について、図24を用いてその動作を述べる。

図24は、図23の音声認識装置において、各構成要素に入力される、または各構成要素から出力される信号の時間波形を示す図である。図24において、スピーカ201からオーディオ信号が拡声された状態で、音声制御用の音声をユーザが発した場合を考える。

【0004】

スピーカ201からオーディオ信号が拡声されない状態で、ユーザが音声を発した場合は、マイクロホン202からは、図24の211に示す極めてS/Nのよい音声信号が出力される。ところが、スピーカ201に図24の212に示すTV番組のオーディオ信号が入力されていると、このスピーカ入力212と相似の図24の213に示す反響信号がマイクロホン202の出力に混入することになる。

【0005】

従って、マイクロホン202からは、ユーザ音声211と反響信号213とが加算された、図24の214に示す、ユーザ音声を認識するには極めてS/Nの悪い信号が出力される。このS/Nの悪いマイクロホン出力214が、音声認識部203に入力されても、音声認識部203で十分な音声認識結果が期待できないのは当然である。

【0006】

そこで、図23の音声認識装置では、スピーカ201からマイクロホン202に回り込む反響信号213を、エコーキャンセラ204内部の適応デジタルフィルタで推定している。そして、エコーキャンセラ204内部の減算回路により、マイクロホン出力214からこの推定反響信号を差し引くことによって反響信号213を完全にキャンセルし、ユーザ音声211だけを抽出している。

【0007】

エコーキャンセラ204には、スピーカ201への入力信号であるスピーカ入力212が与えられる。エコーキャンセラ204内部の適応デジタルフィルタは、このスピーカ入力212の波形と、内部に記憶されたスピーカ201からマイクロホン202までの反響路のインパルスレスポンスとから、反響信号215を推定する。次に、エコーキャンセラ204内部の減算回路が、マイクロホン出力214から、この推定反響信号215を差し引き、それによって、エコーキャンセラ出力216を得ている。

【0008】

このエコーキャンセラ出力216と、ユーザ音声の波形211とを比較すればわかるように、スピーカ201からオーディオが拡声された状態であっても、上記のようなエコーキ

10

20

30

40

50

ャンセラ２０４の反響音打ち消し作用によって、音声認識部２０３が正確な音声認識を行うことが期待できる。

【０００９】

しかしながら、図２３の音声認識装置は、モノラルオーディオ方式のＡＶ機器にしか対応しておらず、複数のスピーカを使用するマルチチャンネルオーディオ方式のＡＶ機器には用いることができないという、大きな欠点を有していた。

【００１０】

図２５は、別の従来のＡＶ機器用音声認識装置の構成を示すブロック図である。図２５の音声認識装置は、２つのスピーカ２２１および２２２を持った２チャンネルオーディオ方式のＡＶ機器に用いられる。

10

図２５において、別の従来の音声認識装置は、マイクロホン２２３と、音声認識部２２４と、２つのエコーキャンセラ２２５および２２６とを備えている。

【００１１】

この従来例は、スピーカ２２１からマイクロホン２２３に回り込む反響音と、スピーカ２２２からマイクロホン２２３に回り込む反響音とを、エコーキャンセラ２２５内部の適応デジタルフィルタと、エコーキャンセラ２２６内部の適応デジタルフィルタとで推定し、マイクロホンの出力信号から、この２つの推定値を差し引くことにより、ユーザ音声だけを抽出するものである。先の図２３の音声認識装置と異なり、図２５の音声認識装置は、ステレオ方式のＡＶ機器に適用できる。

【００１２】

20

しかしながら、図２５の音声認識装置では、オーディオチャンネルの数だけエコーキャンセラを必要とするので、マルチチャンネルオーディオのＡＶ機器に用いる場合には、極めて価格の高い音声認識装置になるという欠点を有していた。

さらに、このように複数のエコーキャンセラを用いたシステムにおいては、エコーキャンセラ間の相互干渉が起こるので、エコーキャンセラの適応動作が極めて不安定であり、適応失敗による反響音の増加や発振などが生じてしまうという大きな欠点も知られていた。

【００１３】

【発明が解決しようとする課題】

ＡＶ機器用の音声認識装置においては、スピーカでオーディオを再生しながら音声認識が行えること、マルチチャンネルオーディオに対応できること、信頼性が高いこと、安価であることが強く要望されている。

30

【００１４】

しかしながら、従来のＡＶ機器用音声認識装置は、前述のように、オーディオチャンネルの数だけエコーキャンセラを必要とするので、マルチチャンネルオーディオ方式のＡＶ機器に用いる場合には、極めて価格が高くなるという課題を有していた。

さらに、エコーキャンセラ間の相互干渉のためにエコーキャンセラの適応動作が極めて不安定となって、適応失敗による反響音の増加や発振などを生じ、その結果、音声認識性能が低下するという別の課題も有していた。

【００１５】

それゆえに、本発明の目的は、スピーカからマルチチャンネル音響が出力されている状態で高精度の音声認識が行え、しかも安価であるような、マルチチャンネルＡＶ機器用の音声認識装置を実現することである。

40

【００１６】

【課題を解決するための手段および発明の効果】

第１の発明は、複数のスピーカを通じてマルチチャンネル音響を出力するＡＶ機器に用いられ、マイクロホンを通じて入力されるユーザ音声を認識して、当該ＡＶ機器に所定の処理動作を行わせるための音声認識装置であって、

複数のスピーカへ向かうマルチチャンネル信号をモノラル化するモノラル化手段、

マイクロホンの出力（以下、マイクロホン出力）と、モノラル化手段の出力（以下、モノラル化信号）とが与えられ、当該モノラル化信号に基づいてマルチチャンネル音響の反響

50

音を推定して、当該マイクロホン出力から当該反響音を除去する１つのエコーキャンセラ、
１つのエコーキャンセラの出力（以下、エコーキャンセラ出力）に基づいてユーザ音声を認識する音声認識手段を備えている。

【００１７】

上記第１の発明では、マルチチャンネル信号をモノラル化して１つのエコーキャンセラに与え、その１つのエコーキャンセラがマイクロホン出力からマルチチャンネル音響の反響音を除去するので、チャンネル数に関わらずただ１つのエコーキャンセラを備えるだけで、スピーカからマルチチャンネル音響が出力されている状態で音声認識が行えるようになる。また、複数のエコーキャンセラを備える場合と異なり、エコーキャンセラ間の相互干渉が生じて音声認識性能が低下することもない。

10

【００１８】

第２の発明は、第１の発明において、複数のスピーカへは、マルチチャンネル信号が入力されることを特徴としている。

【００１９】

上記第２の発明では、複数のスピーカからマルチチャンネル音響が出力されるので、反響音をモノラル化信号で完全に打ち消すことはできない。しかし、マルチチャンネル信号のモノラル度が「１」に近ければ、反響音を概ね打ち消すことができる。少なくとも、マルチチャンネル信号のモノラル度が「０」でない限りは、反響音の一部を打ち消すことができる。

20

ここで、マルチチャンネル信号のモノラル度とは、その信号中に占める、全チャンネルに共通して含まれる成分（モノラル成分）の割合をいい、全チャンネルの信号が互いに全く無相関であれば、モノラル度は「０」、同一であれば、モノラル度は「１」となる。

【００２０】

第３の発明は、第１の発明において、マルチチャンネル信号およびモノラル化信号のいずれかを複数のスピーカへと入力するための切り替え手段をさらに備えている。

【００２１】

上記第３の発明では、複数のスピーカから、マルチチャンネル音響およびモノラル化音響のいずれかを選択的に出力させることができる。

【００２２】

30

第４の発明は、第３の発明において、モノラル化信号とエコーキャンセラ出力とに基づいてユーザ音声を検出する音声検出手段をさらに備え、切り替え手段は、

音声検出手段によってユーザ音声が検出されていない時、マルチチャンネル信号を複数のスピーカへと入力し、

音声検出手段によってユーザ音声が検出されている時、モノラル化信号を複数のスピーカへと入力することを特徴としている。

【００２３】

上記第４の発明では、音声認識を行う必要がない（ユーザ音声が検出されない）場合はマルチチャンネル音響を、音声認識を行う必要がある（ユーザ音声が検出されている）場合はモノラル化音響を出力させるので、十分に高い精度で音声認識が行えるようになる。

40

【００２４】

第５の発明は、第３の発明において、音声認識動作の起動を命令する起動命令手段、音声認識動作の終了を命令する終了命令手段、および起動命令手段および終了命令手段からの命令に応じて、音声認識手段を動作状態および待機状態のいずれかに設定する状態設定手段をさらに備え、切り替え手段は、

状態設定手段によって音声認識手段が待機状態に設定されている時、マルチチャンネル信

50

号を複数のスピーカへと入力し、
状態設定手段によって音声認識手段が動作状態に設定されている時、モノラル化信号を複数のスピーカへと入力することを特徴としている。

【 0 0 2 5 】

上記第 5 の発明では、音声認識手段が待機状態（" O F F " 状態）にある場合はマルチチャンネル音響を、動作状態（" O N " 状態）にある場合はモノラル化音響を出力させるので、十分に高い精度で音声認識が行えるようになる。

【 0 0 2 6 】

第 6 の発明は、第 5 の発明において、
マルチチャンネル信号のモノラル度を判定するモノラル度判定手段、および
マルチチャンネル信号を任意のモノラル度にモノラル化する任意度モノラル化手段をさらに備え、

10

モノラル化手段は、マルチチャンネル信号を完全にモノラル化し、
任意度モノラル化手段は、モノラル度判定手段の判定結果が予め定められたモノラル度よりも低い場合、マルチチャンネル信号を当該予め定められたモノラル度にモノラル化することを特徴としている。

【 0 0 2 7 】

上記第 6 の発明では、マルチチャンネル信号のモノラル度が常に、予め定められたモノラル度以上になるので、音声認識手段が動作状態（" O N " 状態）にある場合でも、立体感をあまり損なうことなく、高い精度で音声認識性能が行えるようになる（すなわち、立体感と音声認識性能とをバランスさせることができる）。

20

【 0 0 2 8 】

第 7 の発明は、第 5 の発明において、
マルチチャンネル信号は、3 チャンネル以上の信号であり、
マルチチャンネル信号を 2 チャンネル化する 2 チャンネル化手段をさらに備え、
モノラル化手段は、2 チャンネル化手段の出力（以下、2 チャンネル化信号）をモノラル化し、

切り替え手段は、マルチチャンネル信号、2 チャンネル化信号およびモノラル化信号のいずれかを複数のスピーカへと入力することを特徴としている。

【 0 0 2 9 】

30

上記第 7 の発明では、複数のスピーカから、マルチチャンネル音響、2 チャンネル化音響およびモノラル化音響のいずれかを選択的に出力させることができる。

【 0 0 3 0 】

第 8 の発明は、第 7 の発明において、
モノラル化信号とエコーキャンセラ出力とに基づいてユーザ音声を検出する音声検出手段をさらに備え、

切り替え手段は、
状態設定手段が音声認識手段を待機状態に設定している時には、マルチチャンネル信号を複数のスピーカへと入力し、

状態設定手段が音声認識手段を動作状態に設定しているが、音声検出手段によってユーザ音声を検出されていない時には、2 チャンネル化信号を複数のスピーカへと入力し、
音声検出手段によってユーザ音声を検出されている時には、モノラル化信号を複数のスピーカへと入力することを特徴としている。

40

【 0 0 3 1 】

上記第 8 の発明では、音声認識手段が待機状態（" O F F " 状態）にある場合はマルチチャンネル音響を、動作状態（" O N " 状態）にあるが音声認識を行う必要がない（ユーザ音声を検出されない）場合はマルチチャンネル音響を、音声認識を行う必要がある（ユーザ音声を検出されている）場合はモノラル化音響を出力させるので、待機状態における立体感をあまり損なうことなく、十分に高い精度で音声認識が行えるようになる。

【 0 0 3 2 】

50

第 9 の発明は、第 5 の発明において、
モノラル化信号とエコーキャンセラ出力とに基づいて、エコーキャンセラにおいて反響音が十分に打ち消されているかを監視する打ち消し監視手段、
モノラル化信号とエコーキャンセラ出力とに基づいてユーザ音声を検出する音声検出手段、および
マルチチャンネル信号を減衰させる減衰手段をさらに備え、
減衰手段は、打ち消し監視手段の監視結果が打ち消し不十分を示している状態において、音声検出手段がユーザ音声を検出した時、マルチチャンネル信号を減衰させることを特徴としている。

【 0 0 3 3 】

上記第 9 の発明では、反響音が十分に打ち消されていない状態でユーザ音声が検出された時、複数のスピーカから出力される音響のレベルを下げることによって、反響音の混入を抑制する。その結果、反響音が十分に打ち消されていない状態での音声認識性能が高まる。

【 0 0 3 4 】

第 1 0 の発明は、第 5 の発明において、
エコーキャンセラは、
複数のスピーカとマイクロホンとの間の反響路のインパルスレスポンスを推定して、当該推定インパルスレスポンスと、モノラル化信号とから反響音を算出する適応デジタルフィルタ、および
マイクロホン出力から、適応デジタルフィルタの出力を減算する減算手段を含んでいる。

【 0 0 3 5 】

上記第 1 0 の発明では、マイクロホン出力からマルチチャンネル音響の反響音を除去して、ユーザ音声だけを音声認識手段に与えることができる。

【 0 0 3 6 】

第 1 1 の発明は、第 1 0 の発明において、切り替え手段によって複数のスピーカへの入力がマルチチャンネル信号からモノラル化信号へと切り替えられる際に、適応デジタルフィルタの適応を促進するためのモノラル適応音を発生する適応音発生手段をさらに備えている。

【 0 0 3 7 】

上記第 1 1 の発明では、スピーカへの入力がマルチチャンネル信号からモノラル化信号に切り替わる際に、複数のスピーカからモノラルの適応音響が出力されるので、もし切り替え直後のモノラル化音響が無音状態であっても、デジタルフィルタが保持しているインパルスレスポンスを強制的に反響路のインパルスレスポンスに適応させることができる。

【 0 0 3 8 】

第 1 2 の発明は、第 1 0 の発明において、
適応デジタルフィルタの適応速度を制御する適応制御手段をさらに備え、
適応制御手段は、モノラル用の速い適応速度と、マルチチャンネル用の遅い適応速度とを持ち、状態設定手段が音声認識手段を動作状態に設定している時には速い適応速度を選択し、待機状態に設定している時には遅い適応速度を選択することを特徴としている。

【 0 0 3 9 】

上記第 1 2 の発明では、エコーキャンセラ内の適応デジタルフィルタの適応速度を、音声認識手段が動作状態に設定されている時には速い速度に、待機状態に設定されている時には遅い速度に制御するので、モノラル、マルチチャンネルそれぞれに好適なエコー打ち消しを行うことができる。

すなわち、スピーカから出力されるのがマルチチャンネル音響の場合は、適応デジタルフィルタから見ると雑音であるステレオ成分が多いので、遅い適応速度とすることによって耐雑音性を高め、一方、モノラル化音響の場合は、ステレオ成分がないので、速い適応速度とすることによって、反響路のインパルスレスポンスの変動への追従性を高めること

10

20

30

40

50

ができる。

その結果、待機状態において優れたエコー打ち消し効果が実現され、動作状態へ移行した直後における音声認識性能が高まる。

【 0 0 4 0 】

第 1 3 の発明は、第 1 2 の発明において、

適応制御手段へは、複数のスピーカに入力される信号がマルチチャンネル信号かモノラル信号かを示す識別信号が与えられ、

適応制御手段は、識別信号がモノラルを示す場合、状態設定手段が音声認識手段を動作状態に設定しているか待機状態に設定しているかに関わらず、速い適応速度を選択することを特徴としている。

10

【 0 0 4 1 】

上記第 1 3 の発明では、複数のスピーカに入力される信号がマルチチャンネル信号かモノラル信号かを識別信号によって判別して、モノラル信号の場合は、状態設定手段が音声認識手段を動作状態に設定しているか待機状態に設定しているかに関わらず、速い適応速度を選択するので、反響路のインパルスレスポンスの変動への追従性が低下することがなく、その結果、待機状態において優れたエコー打ち消し効果が実現され、動作状態へ移行した直後における音声認識性能が高まる。

【 0 0 4 2 】

第 1 4 の発明は、第 1 0 の発明において、

マルチチャンネル信号のモノラル度を判定するモノラル度判定手段、および

20

モノラル度判定手段の判定結果に基づいて、適応デジタルフィルタの適応速度を制御する適応制御手段をさらに備えている。

【 0 0 4 3 】

上記第 1 4 の発明では、マルチチャンネル信号のモノラル度に基づいて、適応デジタルフィルタの適応速度を制御するので、様々なモノラル度を持ったマルチチャンネル信号に好適なエコー打ち消しを行うことができる。

すなわち、モノラル度が低い場合、適応速度を遅くして、耐雑音性を高める。

一方、モノラル度が高い場合、適応デジタルフィルタから見ると雑音であるステレオ成分が少ないので、耐雑音性はあまり必要でない。よって、下記第 1 5 の発明のように、適応速度を速くすることによって、反響路のインパルスレスポンスの変動への追従性を高めることができる。その結果、特にモノラル度が高い場合に、優れたエコー打ち消し効果が実現でき、動作状態へ移行した直後における音声認識性能が高まる。

30

【 0 0 4 4 】

第 1 5 の発明は、第 1 4 の発明において、適応制御手段は、マルチチャンネル信号のモノラル度が高いほど、適応デジタルフィルタの適応速度を速くすることを特徴としている。

【 0 0 4 5 】

第 1 6 の発明は、第 1 0 の発明において、

不揮発メモリをさらに備え、

不揮発メモリは、

40

電源が " O F F " される時、適応デジタルフィルタによって推定されたインパルスレスポンスを取得して記憶し、

電源が " O N " される時、記憶している電源 " O F F " 時の推定インパルスレスポンスを適応デジタルフィルタに与え、

適応デジタルフィルタは、不揮発メモリから与えられた電源 " O F F " 時の推定インパルスレスポンスを初期値として、インパルスレスポンスの推定を開始することを特徴としている。

【 0 0 4 6 】

上記第 1 6 の発明では、電源 " O F F " 時の推定インパルスレスポンスを記憶しておき、電源 " O N " 時、それを初期値としてインパルスレスポンスの推定を開始するので、 " 0

50

”を初期値とする場合と比べ、電源”ON”直後の推定誤差が小さくなり、その結果、音声認識性能が高まる。

【0047】

第17の発明は、第5の発明において、モノラル化信号とエコーキャンセラ出力とに基づいてユーザ音声を検出する音声検出手段をさらに備え、起動命令手段は、ボタンが押されると、状態設定手段へ起動命令を発するようなボタンスイッチであり、終了命令手段は、音声検出手段がユーザ音声を検出しない状態が予め定められた時間以上継続すると、状態設定手段へ終了命令を発する時限スイッチであることを特徴としている。

10

【0048】

上記第17の発明では、音声認識動作を自動的に終了できる。

【0049】

第18の発明は、第5の発明において、モノラル化信号とエコーキャンセラ出力とに基づいてユーザ音声を検出する音声検出手段をさらに備え、起動命令手段は、音声検出手段がユーザ音声を検出すると、状態設定手段へ起動命令を発する音声スイッチであり、終了命令手段は、音声検出手段がユーザ音声を検出しない状態が予め定められた時間以上継続すると、状態設定手段へ終了命令を発する時限スイッチであることを特徴としている。

20

【0050】

上記第18の発明では、音声認識動作を自動的に開始・終了できる。

【0051】

【発明の実施の形態】

以下、本発明の実施の形態について、図面を参照しながら説明する。

最初に、本発明が用いられるAV機器について説明しておく。

図1は、本発明が用いられるAV機器の構成の一例を示すブロック図である。

図1に示されるAV機器は、テレビジョン放送を受信するためのテレビジョン受像機である。ここでいうテレビジョン放送では、マルチチャンネル(2チャンネルも含む；以下同様)の音響方式が採用されているとする。

30

【0052】

図1において、AV機器は、アンテナ1と、受信部2と、AV処理部3と、コントローラ4と、コントロールパネル5と、マイクロホン6と、音声認識装置7と、ディスプレイユニット8と、スピーカユニット9とを備えている。

【0053】

アンテナ1は、放送局から送信された電波を捉えて電気信号に変換する。受信部2は、アンテナ1から出力される電気信号から、特定の周波数帯に含まれる信号を抽出する。AV処理部3は、受信部2から出力される信号を処理して、映像信号と、マルチチャンネルの音響信号(以下、マルチチャンネル信号)とを出力する。

40

【0054】

コントローラ4は、コントロールパネル5または音声認識装置7からのコントロール信号を受け、受信部2および/またはAV処理部3に、例えば受信チャンネルの切り換えや音量の増減、主電源の”ON”/”OFF”等の予め決められた処理を実行させる。ディスプレイユニット8は、ディスプレイを含み、AV処理部3からの映像信号を受けて映像を表示する。スピーカユニット9は、複数のスピーカ(9a, 9b, ...)を含み、AV処理部3からのマルチチャンネル信号を受けてマルチチャンネル音響を出力する。

【0055】

コントロールパネル5(受像機本体に設けられても、リモコンに設けられてもよい)は、

50

ボタン等によって構成され、ユーザのボタン操作と対応するコントロール信号を生成する。マイクロホン 6 は、ユーザが発した音声を電気信号に変換する。音声認識装置 7 は、マイクロホン 6 から出力される電気信号を受け、ユーザの音声と対応するコントロール信号を生成する。

【 0 0 5 6 】

ここで、上記受信部 2 から出力される信号は、アナログ信号であっても、デジタル信号であってもよい。前者の場合、A V 処理部 3 は、受信部 2 から出力される信号をアナログ的に処理する回路によって構成される。後者の場合、A V 処理部 3 は、受信部 2 から出力される信号をデジタル的に処理する回路によって構成される。

【 0 0 5 7 】

上記のように構成されたテレビジョン受像機では、アンテナ 1 が、放送局から送信された電波を捉えて電気信号に変換し、受信部 2 は、その電気信号から、特定の周波数帯の信号を抽出する。次いで、A V 処理部 3 は、受信部 2 から出力される信号を処理して、映像信号とマルチチャンネル信号とを出力する。A V 処理部 3 から出力された映像信号は、ディスプレイユニット 8 に与えられ、ディスプレイに映像が表示される。一方、マルチチャンネル信号は、スピーカユニット 9 に与えられ、複数のスピーカからマルチチャンネル音響が出力される。

【 0 0 5 8 】

ユーザは、コントロールパネル 5 を操作することにより、テレビジョン受像機に、受信チャンネルの切換等を行わせることができる。すなわち、コントロールパネル 5 がユーザのボタン操作と対応するコントロール信号を生成し、コントローラ 4 は、そのコントロール信号を受け、受信部 2 および / または A V 処理部 3 に、受信チャンネルの切り換え等を実行させる。

【 0 0 5 9 】

また、ユーザは、マイクロホン 6 を通じて音声を入力をすることにより、テレビジョン受像機に、受信チャンネルの切り換え等を行わせることができる。すなわち、音声認識装置 7 がユーザの音声と対応するコントロール信号を生成し、コントローラ 4 は、そのコントロール信号を受け、受信部 2 および / または A V 処理部 3 に、受信チャンネルの切り換え等を実行させる。

【 0 0 6 0 】

なお、以上では、本発明が用いられる A V 機器の一例として、マルチチャンネル音響を出力するテレビジョン受像機について説明したが、本発明は、テレビジョン受像機に限らず、例えば、マルチチャンネル音響を出力するラジオ受信機に用いてもよい。あるいは、マルチチャンネル信号が記された C D , D V D 等のメディアを再生するための再生機、アンプ、およびスピーカユニット 9 等からなるマルチチャンネル・オーディオシステムなど、マルチチャンネル音響を出力する機能を持つあらゆる機器、システムに用いることができる。

【 0 0 6 1 】

(第 1 の実施形態)

図 2 は、本発明の第 1 の実施形態に係る A V 機器用音声認識装置の構成を示すブロック図である。図 2 の音声認識装置 7 は、図 1 の A V 機器に設けられた音声認識装置 7 と対応している。ただし、本実施形態では、A V 機器において、A V 処理部 3 から 2 チャンネル信号が出力され、スピーカユニット 9 に含まれる 2 つのスピーカ 9 a および 9 b を通じて 2 チャンネル音響が出力されるものとする。

【 0 0 6 2 】

図 2 において、音声認識装置 7 は、モノラル化部 1 3 と、1 個のエコーキャンセラ 1 4 と、音声認識部 1 5 とを備えている。スピーカ 9 a および 9 b へと入力される信号は、図 1 の A V 処理部 3 から出力される 2 チャンネル信号である。

【 0 0 6 3 】

スピーカ 9 a および 9 b へ向かう 2 チャンネル信号が分岐されてモノラル化部 1 3 に入力

10

20

30

40

50

され、モノラル化部 13 は、その 2 チャンネル信号をモノラル化する。

マイクロホン 6 から出力される信号（以下、マイクロホン出力）と、モノラル化部 13 から出力される信号（以下、モノラル化信号）とがエコーキャンセラ 14 へと与えられ、エコーキャンセラ 14 は、そのマイクロホン出力からユーザの音声と対応する信号（以下、ユーザ音声）だけを抽出する。

【0064】

ここで、エコーキャンセラ 14 の動作原理について、簡単に説明する。エコーキャンセラ 14 は、適応デジタルフィルタ 14 a と、減算回路 14 b とを含む。マイクロホン出力には、ユーザ音声に加え、スピーカ 9 a および 9 b から出力された音響が屋内を反響してマイクロホン 6 へと回り込んだ結果生じる信号（以下、反響信号）が含まれている。

10

【0065】

適応デジタルフィルタ 14 a へは、モノラル化信号が入力されると共に、減算回路 14 b から出力される信号がフィードバックされ、適応デジタルフィルタ 14 a は、それら 2 つの信号に基づいて反響信号を推定する。こうして得られた推定反響信号と、マイクロホン出力とが減算回路 14 b へと与えられ、減算回路 14 b は、マイクロホン出力から推定反響信号を減算する。これにより、エコーキャンセラ 14 からは、反響信号の除去されたユーザ音声出力される。

【0066】

音声認識部 15 は、エコーキャンセラ 14 からのユーザ音声を認識して、その音声を示すコントロール信号を生成する。こうして生成されたコントロール信号は、図 1 のコントローラ 4 に伝達され、コントローラ 4 が受信部 2 や A/V 処理部 3 を制御することによって、テレビジョン受像機において、受信チャンネルの切り換え等の処理が実行される。

20

【0067】

図 3 は、図 1 の音声認識装置 7 のハードウェア的な構成を示すブロック図である。図 3 において、音声認識装置 7 は、CPU 10 と、RAM 11 と、ROM 12 とを備えている。ROM 12 には、所定のプログラムが予め格納されている。

このプログラムには、(a) 2 チャンネル信号をモノラル化するためのアルゴリズムや、(b) マイクロホン出力から反響信号を除去するためのアルゴリズム、(c) ユーザ音声を認識してコントロール信号を生成するためのアルゴリズム等が記述されている。CPU 10 は、RAM 11 を作業領域として利用しつつ、上記のプログラムに従って動作する。これによって、図 2 に示される各ブロックの機能が実現される。なお、各ブロックの機能をソフト的に実現する代わりに、各々専用のハード回路によって実現することもできる。

30

【0068】

以上のように構成された A/V 機器用音声認識装置 7 について、以下、図 4 を利用してその動作を述べる。図 4 は、図 2 の音声認識装置 7 において、各構成要素に入力される、または各構成要素から出力される信号の時間波形を示す図である。

まず、スピーカ 9 a とスピーカ 9 b とに図 4 の 21 に示す左スピーカ入力と図 4 の 22 に示す右スピーカ入力とが入力されている状態で、ユーザが図 4 の 23 に示す音声を発した場合を考える。この時、マイクロホン 6 からは、左スピーカ入力 21 の反響音と右スピーカ入力 22 の反響音とユーザ音声 23 とが加算された図 4 の 24 に示すマイクロホン出力信号が出力される。一方、左スピーカ入力 21 と右スピーカ入力 22 とは、モノラル化部 13 にも入力され、ここで加算されて、図 4 の 25 に示すモノラル化信号が得られる。

40

【0069】

このモノラル化信号 25 がエコーキャンセラ 14 に入力され、エコーキャンセラ 14 は、このモノラル化信号 25 と内部に記憶した推定インパルスレスポンスとから、図 4 の 26 に示す推定反響信号を推定する。エコーキャンセラ 14 の内部では、マイクロホン出力信号 24 から、推定した推定反響信号 26 を差し引きくことにより、図 4 の 27 に示すエコーキャンセラ出力信号を得る。この信号が音声認識部 15 に入力される。エコーキャンセラ出力信号 27 とユーザ音声 23 とマイクロホン出力信号 24 とを比較すれば、反響信号がかなり効果的に打ち消されていることがわかる。

50

【0070】

次に、なぜステレオ信号（以下では、2チャンネル信号を適宜、ステレオ信号と呼ぶ）が1個のエコーキャンセラ14で打ち消せるかを説明する。

右チャンネル用のスピーカ9aからマイクロホン6までの伝達特性（インパルスレスポンス）を H_r 、左チャンネル用のスピーカ9bからマイクロホン6までの伝達特性を H_l とし、右チャンネル信号を S_r 、左チャンネル信号を S_l とすると、マイクロホン6の出力に混入する反響信号 S_e は、 $S_e = (S_r * H_r + S_l * H_l)$ となる。

【0071】

この時、もし、左右の伝達特性がほぼ等しく $H_r = H_l (= H)$ なら、 $S_e = (S_r + S_l) * H$ となり、もし、左右のチャンネルの信号がほぼ等しく $S_r = S_l = S$ なら、 $S_e = S * (H_r + H_l)$ となる。従って、どちらかの仮定が成り立つ場合には、1個のエコーキャンセラ14でも打ち消しが可能であることがわかる。

10

【0072】

伝達特性 H_r および H_l を決める最大の要因は、スピーカ9aおよび9bとマイクロホン6との間の距離および部屋の反射構造であるが、実際の聴取状態では、音声認識用のマイクロホン6と、右チャンネル用のスピーカ9a、左チャンネル用のスピーカ9bとの間の各距離は、例えばマイクロホン6をユーザの手元に置く場合には、当然、ほぼ等距離となる。また、TVの上に設置する場合でも、TVの中央に設置すれば等距離となる。さらに、部屋の反響構造も当然ほぼ同一である。

20

【0073】

高域周波数では、波長が短いため僅かな距離差によって位相逆転が生じるので、たとえばほぼ等距離といえども位相を含めた伝達特性の一致は不十分である。しかし、低中域周波数では、かなりよく伝達特性が一致している場合が多いので、 $H_r = H_l$ の仮定が成立し、1個のエコーキャンセラ14でもある程度のキャンセル効果が期待できることになる。

【0074】

さらに、実際のTV番組などでの音づくりでは、センター定位音（モノラル成分）を比較的高いレベルで左右チャンネル均等にミキシングし、このモノラル成分に、左右に定位させる音（ステレオ成分）を比較的低レベルでミキシングしている場合が多い。すなわち、センター音源主体の音づくりがなされており、先の21と22で示した左右のスピーカ入力のかかなりの部分がモノラル成分である。

30

このようなセンター音源主体のオーディオ信号の場合には、 $S_r = S_l$ の仮定が成り立ち、1個のエコーキャンセラ14を用いたシステムでも、反響音を効果的にキャンセルできることになる。

以上の理由により、実際のTV視聴状態においては、図2のような構成の音声認識装置7でも、かなりのエコーキャンセル効果が得られることが確かめられている。

【0075】

以上のように、本実施形態によれば、1個のエコーキャンセラのみでステレオソース（2チャンネル信号）に対応できるため、安価なAV機器用音声認識装置が実現できる。また、1個のエコーキャンセラしか使用しないため、エコーキャンセラ間の相互干渉がなく、安定した動作が保証できるという、実用上極めて重要な効果が得られる。

40

【0076】

なお、上記第1の実施形態（および下記第2～4、6～13の実施形態）では、図1のAV機器において、AV処理部3から2チャンネル信号が出力され、スピーカユニット9を通じて2チャンネル音響が拡声されとしたが、代わりに、AV処理部3から4チャンネル信号、6チャンネル信号等のマルチチャンネル信号が出力され、スピーカユニット9を通じて4チャンネル音響、6チャンネル音響等のマルチチャンネル音響が出力されてもよい。この場合、図2のモノラル化部13がマルチチャンネル信号のモノラル化を行うように、ROM12のプログラムの記述（あるいは専用ハード回路の構成）に一部変更を加えればよい。

その場合、モノラル化部13は、全チャンネルの信号を加算してもよく、あるいは、前方

50

左右中央といった主チャンネル信号だけを加算してもよい。また、加算する際に、各チャンネルを均等に加算するのではなく、重み付けをして加算してもよい。

【 0 0 7 7 】

(第 2 の実施形態)

図 5 は、本発明の第 2 の実施形態に係る A V 機器用音声認識装置の構成を示すブロック図である。図 5 の音声認識装置 7 は、図 1 の A V 機器に設けられた音声認識装置 7 と対応している。ただし、本実施形態では、A V 機器において、A V 処理部 3 から 2 チャンネル信号が出力され、スピーカユニット 9 に含まれる 2 つのスピーカ 9 a および 9 b を通じて 2 チャンネル音響が出力されるものとする。

【 0 0 7 8 】

図 5 において、音声認識装置 7 は、モノラル化部 3 3 と、1 個のエコーキャンセラ 3 4 と、音声認識部 3 5 と、音声検出部 3 7 と、切り替え部 3 6 とを備えている。すなわち、図 5 の音声認識装置 7 は、図 2 の音声認識装置 7 (第 1 の実施形態) に、音声検出部 3 7 と、切り替え部 3 6 とを追加したものである。スピーカ 9 a および 9 b へと入力される信号は、図 1 の A V 処理部 3 から出力される 2 チャンネル信号である。

【 0 0 7 9 】

スピーカ 9 a および 9 b へ向かう 2 チャンネル信号が分岐されてモノラル化部 3 3 に入力され、モノラル化部 3 3 は、その 2 チャンネル信号をモノラル化する。

マイクロホン 6 から出力される信号 (マイクロホン出力) と、モノラル化部 3 3 から出力される信号 (モノラル化信号) とがエコーキャンセラ 3 4 へと与えられ、エコーキャンセラ 3 4 は、そのマイクロホン出力からユーザの音声と対応する信号 (以下、ユーザ音声) だけを抽出する。なお、エコーキャンセラ 3 4 の動作原理については、第 1 の実施形態で説明した。

【 0 0 8 0 】

音声検出部 3 7 へは、モノラル化部 3 3 の出力 (モノラル化信号) と、エコーキャンセラ 3 4 の出力 (ユーザ音声) とが与えられ、音声検出部 3 7 は、両出力のレベル比に基づいて、ユーザ音声を検出する。切り替え部 3 6 は、音声検出部 3 7 がユーザ音声を検出すると、スピーカ 9 a および 9 b への入力を、(非検出時における) 2 チャンネル信号から、モノラル化信号に切り替える。また、音声検出部 3 7 がユーザ音声を検出している状態から検出しない状態へ変化すると、スピーカ 9 a および 9 b への入力を、(検出時における) モノラル化信号から、2 チャンネル信号に切り替える。

【 0 0 8 1 】

音声認識部 3 5 は、音声検出部 3 7 がユーザ音声を検出すると、音声認識動作を開始する。すなわち、エコーキャンセラ 3 4 からのユーザ音声を認識して、その音声が表示コントロール信号を生成する。こうして生成されたコントロール信号は、図 1 のコントローラ 4 に伝達され、コントローラ 4 が受信部 2 や A V 処理部 3 を制御することによって、テレビジョン受像機において、受信チャンネルの切り換え等の処理が実行される。

【 0 0 8 2 】

図 5 の音声認識装置 7 のハードウェア的な構成は、図 3 と同様である。図 3 において、ROM 1 2 には、所定のプログラムが予め格納されている。このプログラムには、第 1 の実施形態で述べた (a) ~ (c) のアルゴリズムに加え、さらに (d) ユーザ音声を検出するためのアルゴリズム、および (e) スピーカ 9 a および 9 b への入力信号を切り替えるためのアルゴリズムが記述されている。CPU 1 0 は、RAM 1 1 を作業領域として利用しつつ、上記のプログラムに従って動作する。これによって、図 5 に示される各ブロックの機能が実現される。なお、各ブロックの機能をソフト的に実現する代わりに、各々専用のハード回路によって実現することもできる。

【 0 0 8 3 】

以上のように構成された A V 機器用音声認識装置 7 について、以下、図 6 を利用してその動作を述べる。図 6 は、図 5 の音声認識装置 7 において、各構成要素に入力される、または各構成要素から出力される信号の時間波形を示す図である。

まず、先の第1実施形態で説明したと同様、スピーカ9aとスピーカ9bとに、図6の41に示す左スピーカ入力と図6の42に示す右スピーカ入力とが入力されている状態で、ユーザが図6の43に示す音声を発した場合を考える。この時、モノラル化部33からは、図6の44の信号が出力される。図3の音声認識装置7では、音声検出部37がユーザが音声を発したか否かを判断し、ユーザが音声を発していない図6の43に示したA～BとC～Dの期間では、切り替え部36は、スピーカ9aおよび9bへの入力をステレオ信号側に切り替え、ユーザが音声を発しているB～Cの期間では、モノラル信号側に切り替えている。この時、マイクロホン6から出力される信号を、図6の45に示す。

【0084】

エコーキャンセラ34には、常に図6のモノラル化信号44が入力され、エコーキャンセラ34は、このモノラル化信号44と、内部に記憶した推定インパルスレスポンスとから、図6の46に示す推定反響信号を推定する。エコーキャンセラ34の内部では、マイクロホン出力信号45から推定反響信号46を差し引くことにより、図6の47に示すエコーキャンセラ出力信号を得る。

【0085】

図5の音声認識装置7では、音声検出部37が、モノラル化信号44とエコーキャンセラ出力信号47とのレベル比を監視し、モノラル化信号44のレベルと反響路の伝達特性とから予想される反響信号のレベル以上に、エコーキャンセラ出力信号47のレベルが上昇した時に、ユーザが音声を発したと判断し、切り替え部36は、スピーカ9aおよび9bへの入力をモノラル化信号44に切り替える。スピーカ9aおよび9bへの入力を($S_r + S_l$)なるモノラル化信号に切り替えた場合においては、反響信号 S_e は、 $S_e = (S_r + S_l) * (H_r + H_l)$ となり、原理的に、1個のエコーキャンセラ34により反響信号を完全に消去できることとなる。先の図2の構成では、 S_r 、 S_l の仮定が成り立たないステレオ成分の強いオーディオ信号の場合には、当然、エコーキャンセラ34の打ち消し効果が損なわれ、音声認識部35に入力される音声に反響信号が混入して音声認識の性能を悪化させていたが、図5の構成では、この場合においても完全に反響信号を打ち消すことができ、音声認識部35で精度の高い音声認識が可能となる。

【0086】

図6の48に、図5の音声認識装置7におけるエコーキャンセラ出力信号47中に含まれる反響信号を抜き出して示し、図6の49に、図2の構成における図2のエコーキャンセラ出力信号27中に含まれる反響信号を抜き出して示す。この48と49を比較すると、本実施の形態では、B～Cのユーザ音声が入力されている期間において反響信号がより効果的に打ち消され、音声認識のためのS/Nが大幅に向上していることがわかる。

【0087】

以上のように、本実施形態によれば、通常はステレオ再生で、ユーザが音声を発した時のみモノラル再生に切り替えることにより、第1の実施形態よりさらにS/Nのよいユーザ音声を抽出でき、認識性能の向上が図れる。

【0088】

(第3の実施形態)

図7は、本発明の第3の実施形態に係るAV機器用音声認識装置の構成を示すブロック図である。図7の音声認識装置7は、図1のAV機器に設けられた音声認識装置7と対応している。ただし、本実施形態では、AV機器において、AV処理部3から2チャンネル信号が出力され、スピーカユニット9に含まれる2つのスピーカ9aおよび9bを通じて2チャンネル音響が出力されるものとする。

【0089】

図7において、音声認識装置7は、モノラル化部53と、1個のエコーキャンセラ54と、音声認識部55と、起動命令部581と、終了命令部582と、状態設定部57と、切り替え部56とを備えている。すなわち、図7の音声認識装置7は、図2の音声認識装置7(第1の実施形態)に、起動命令部581と、終了命令部582と、状態設定部57と、切り替え部56とを追加したものである。スピーカ9aおよび9bへと入力される信号

10

20

30

40

50

は、図 1 の A V 処理部 3 から出力される 2 チャンネル信号である。

【 0 0 9 0 】

スピーカ 9 a および 9 b へ向かう 2 チャンネル信号が分岐されてモノラル化部 5 3 に入力され、モノラル化部 5 3 は、その 2 チャンネル信号をモノラル化する。

マイクロホン 6 から出力される信号（マイクロホン出力）と、モノラル化部 5 3 から出力される信号（モノラル化信号）とがエコーキャンセラ 5 4 へと与えられ、エコーキャンセラ 5 4 は、そのマイクロホン出力からユーザの音声と対応する信号（以下、ユーザ音声）だけを抽出する。なお、エコーキャンセラ 5 4 の動作原理については、第 1 の実施形態で説明した。

【 0 0 9 1 】

起動司令部 5 8 1 は、音声認識動作の起動を命令する。終了司令部 5 8 2 は、音声認識動作の終了を命令する。状態設定部 5 7 は、起動司令部 5 8 1 , 終了司令部 5 8 2 からの命令を受け、音声認識部 5 5 の動作状態を設定（つまり音声認識動作を " O N " / " O F F " ）する。切り替え部 5 6 は、状態設定部 5 7 が音声認識動作を " O N " 状態に設定すると、スピーカ 9 a および 9 b への入力を、（ " O F F " 状態における）2 チャンネル信号から、モノラル化信号に切り替える。また、 " O F F " 状態に設定すると、スピーカ 9 a および 9 b への入力を、（ " O N " 状態における）モノラル化信号から、2 チャンネル信号に切り替える。

【 0 0 9 2 】

音声認識部 5 5 は、状態設定部 5 7 の設定に従い、音声認識を実行 / 終了する。すなわち、エコーキャンセラ 5 4 からのユーザ音声を認識して、その音声が表示コントロール信号を生成する。こうして生成されたコントロール信号は、図 1 のコントローラ 4 に伝達され、コントローラ 4 が受信部 2 や A V 処理部 3 を制御することによって、テレビジョン受像機において、受信チャンネルの切り換え等の処理が実行される。

【 0 0 9 3 】

図 7 の音声認識装置 7 のハードウェア的な構成は、図 3 と同様である。図 3 において、R O M 1 2 には、所定のプログラムが予め格納されている。このプログラムには、第 1 の実施形態で述べた（ a ） ~ （ c ）のアルゴリズムと、第 2 の実施形態で述べた（ e ）のアルゴリズムに加え、さらに（ f ）音声認識部 5 5 の動作状態を設定するためのアルゴリズムが記述されている。C P U 1 0 は、R A M 1 1 を作業領域として利用しつつ、上記のプログラムに従って動作する。これによって、図 7 に示される各ブロックの機能が実現される。

【 0 0 9 4 】

なお、起動司令部 5 8 1 , 終了司令部 5 8 2 は、図 1 のコントローパネルを構成するボタンによって実現される。また、起動司令部 5 8 1 , 終了司令部 5 8 2 以外の各ブロックの機能をソフト的に実現する代わりに、各々専用のハード回路によって実現することもできる。

【 0 0 9 5 】

以上のように構成された A V 機器用音声認識装置 7 について、以下、その動作を述べる。図 7 の音声認識装置 7 は、図 5 の音声認識装置 7 において、音声検出部 3 7 が行っていた切り替え部 5 6 の制御を、起動司令部 5 8 1 および終了司令部 5 8 2 と、状態設定部 5 7 とで行う構成である。音声認識機能を利用しようとする場合は、まず、起動司令部 5 8 1 により音声認識機能の起動信号が状態設定部 5 7 に送られ、状態設定部 5 7 は、切り替え部 5 6 を制御してスピーカ 9 a およびスピーカ 9 b への入力信号をステレオ信号からモノラル化信号へ切り替え、音声認識の状態をスピーカ 9 a および 9 b にステレオ信号が入力された待機状態から、スピーカ 9 a および 9 b にモノラル化信号が入力された動作状態に移行させる。

【 0 0 9 6 】

動作状態においては、ステレオ感は損なわれるがオーディオ信号の聞き取りには大きな問題はない。すなわち、動作状態においては、反響信号の打ち消し効果は最良の状態にあり

10

20

30

40

50

、精度の高い音声認識が期待できる一方、ステレオ感が損なわれるため、音声認識機能の利用が終了すると同時に、終了命令部 582 により音声認識機能の終了信号が状態設定部 57 に送られ、状態設定部 57 は、切り替え部 56 を制御してスピーカ 9a およびスピーカ 9b への入力信号をモノラル化信号からステレオ信号へ切り替え、音声認識の状態をスピーカ 9a および 9b にモノラル化信号が入力された動作状態から、スピーカ 9a および 9b にステレオ信号が入力された待機状態に移行させる。

【0097】

図 8 は、図 7 の音声認識装置 7 において、各構成要素に入力される、または各構成要素から出力される信号の時間波形を示す図である。図 8 の 61 に、動作状態における音声認識部 55 への入力信号を示し、62 に、その信号中に含まれる反響信号を示す。この図 8 の 61, 62 の信号と、先の図 6 の 47, 48 の信号とを比較すれば、図 7 の音声認識装置 7 では、図 5 の音声認識装置 7 よりも語頭や語尾部分の S/N が大幅に改善されることがわかる。先の図 5 の構成では、音声検出に数 10 msec の検出時間を要するため、数 10 msec の語頭部分での S/N が悪く、語頭にある子音を認識しにくいという欠点があったが、図 7 の構成では、この欠点が完全に解消される。

10

【0098】

以上のように、本実施形態によれば、通常はステレオ再生で、音声認識機能を必要とした時のみモノラル再生に切り替えることにより、第 2 の実施形態よりさらに S/N のよいユーザ音声を抽出でき、認識性能のさらなる向上が図れる。

【0099】

20

(第 4 の実施形態)

図 9 は、本発明の第 4 の実施形態に係る AV 機器用音声認識装置の構成を示すブロック図である。図 9 の音声認識装置 7 は、図 1 の AV 機器に設けられた音声認識装置 7 と対応している。ただし、本実施形態では、AV 機器において、AV 処理部 3 から 2 チャンネル信号が出力され、スピーカユニット 9 に含まれる 2 つのスピーカ 9a および 9b を通じて 2 チャンネル音響が出力されるものとする。

【0100】

図 9 において、音声認識装置 7 は、完全モノラル化部 75 と、モノラル度判定部 76 と、任意度モノラル化部 77 と、1 個のエコーキャンセラ 73 と、音声認識部 74 と、起動命令部 792 と、終了命令部 793 と、状態設定部 791 と、切り替え部 78 とを備えている。すなわち、図 9 の音声認識装置 7 は、図 7 の音声認識装置 7 (第 3 の実施形態) に、モノラル度判定部 76 と、任意度モノラル化部 77 とを追加したものである(なお、完全モノラル化部 75 は、任意度モノラル化部 77 と区別するために「完全」と呼んでいるが、図 7 のモノラル化部 53 と同様のものである)。スピーカ 9a および 9b へと入力される信号は、図 1 の AV 処理部 3 から出力される 2 チャンネル信号である。

30

【0101】

スピーカ 9a および 9b へ向かう 2 チャンネル信号が分岐されて完全モノラル化部 75 に入力され、完全モノラル化部 75 は、その 2 チャンネル信号を完全にモノラル化する。さらに、スピーカ 9a および 9b へ向かう 2 チャンネル信号が分岐されてモノラル度判定部 76 および任意度モノラル化部 77 に入力され、モノラル度判定部 76 は、その 2 チャンネル信号のモノラル度を判定する。任意度モノラル化部 77 は、モノラル度判定部 76 の判定結果を受けて、その 2 チャンネル信号を任意の程度にモノラル化する。

40

【0102】

すなわち、任意度モノラル化部 77 は、その 2 チャンネル信号のモノラル度に応じて、その 2 チャンネル信号のモノラル度を高める処理を施す。そのために、任意度モノラル化部 77 は、モノラル度を高める処理をどの強度で行うべきかをモノラル度をもとに決定するための関数(処理強度決定特性; 図 12 (A) に参照番号 101 で示される)を記憶している。

【0103】

ここで、上記 2 チャンネル信号のモノラル度とは、その信号中に占める、両チャンネルに

50

共通して含まれる信号成分（モノラル成分）の割合をいい、両チャンネルの信号が互いに全く無相関であれば、モノラル度は「0」、同一であれば、モノラル度は「1」となる。

【0104】

マイクロホン6から出力される信号（マイクロホン出力）と、完全モノラル化部75から出力される信号（完全モノラル化信号）とがエコーキャンセラ73へと与えられ、エコーキャンセラ73は、そのマイクロホン出力からユーザの音声と対応する信号（以下、ユーザ音声）だけを抽出する。なお、エコーキャンセラ73の動作原理については、第1の実施形態で説明した。

【0105】

起動命令部792は、音声認識動作の起動を命令する。終了命令部793は、音声認識動作の終了を命令する。状態設定部791は、起動命令部792、終了命令部793からの命令を受け、音声認識部74の動作状態を設定（つまり音声認識動作を「ON」/「OFF」）する。

【0106】

任意度モノラル化部77から出力される信号（以下、任意度モノラル化信号）と、図1のAV処理部3からの2チャンネル信号とが切り替え部78へと与えられ、切り替え部78は、状態設定部791が音声認識動作を「ON」状態に設定すると、スピーカ9aおよび9bへの入力を、（「OFF」状態における）2チャンネル信号から、任意度モノラル化信号に切り替える。また、「OFF」状態に設定すると、スピーカ9aおよび9bへの入力を、（「ON」状態における）任意度モノラル化信号から、2チャンネル信号に切り替える。

【0107】

音声認識部74は、状態設定部791の設定に従い、音声認識を実行/終了する。すなわち、エコーキャンセラ73からのユーザ音声を認識して、その音声が表示コントロール信号を生成する。こうして生成されたコントロール信号は、図1のコントローラ4に伝達され、コントローラ4が受信部2やAV処理部3を制御することによって、テレビジョン受像機において、受信チャンネルの切り換え等の処理が実行される。

【0108】

図9の音声認識装置7のハードウェア的な構成は、図3と同様である。図3において、ROM12には、所定のプログラムが予め格納されている。このプログラムには、第1の実施形態で述べた（a）～（c）のアルゴリズムと、第2の実施形態で述べた（e）のアルゴリズムと、第3の実施形態で述べた（f）のアルゴリズムとに加え、さらに（g）2チャンネル信号のモノラル度を判定するためのアルゴリズム、および（h）2チャンネル信号を任意の程度にモノラル化するためのアルゴリズムが記述されている。CPU10は、RAM11を作業領域として利用しつつ、上記のプログラムに従って動作する。これによって、図9に示される各ブロックの機能が実現される。

【0109】

なお、起動命令部792、終了命令部793は、図1のコントローパネルを構成するボタンによって実現される。また、起動命令部792、終了命令部793以外の各ブロックの機能をソフト的に実現する代わりに、各々専用のハード回路によって実現することもできる。

【0110】

以上のように構成されたAV機器用音声認識装置7について、以下、その動作を述べる。図7の音声認識装置7では、音声認識の動作状態（「ON」状態）においては、ステレオ信号を完全にモノラル化してからスピーカ9aおよび9bで再生しているため、ステレオ感がまったく無いという欠点があった。一方、音声認識の動作状態でもステレオ再生を行った図2の音声認識装置7では、Sr、Slの仮定が大きく崩れるモノラル度の低いオーディオ信号が入力している時に、エコーキャンセラ14のエコー打ち消し量が大きく悪化するという欠点を有していた。ただし、前述したように、通常のTVなどのステレオ番組にあっては、エコーキャンセラ14の打ち消し量を大きく悪化させるほど、相関の低いオ

10

20

30

40

50

オーディオ信号が左右のチャンネルにミキシングされている場面は少なく、左右均等にミキシングしたセンター音源に比較的弱いレベルで左右の音源をミキシングしている場面がほとんどである。このため、たとえステレオ放送であっても、左右のチャンネル間の相関が極めて強い $S_r = S_l$ の仮定が成り立つ時間が多い。問題は、相関の低い時間をどう乗り切るかである。

【0111】

そこで、本実施形態では、エコーキャンセラ 73 へは常に、完全モノラル化部 75 で完全にモノラル化信号を入力しているが、スピーカ 9a および 9b へは、音声認識の待機状態（"OFF" 状態）ではステレオ信号を、動作状態では任意度モノラル化部 77 の出力を入力する。モノラル度判定部 76 が信号のモノラル度を監視し、モノラル度が低いと判定された時にのみ、任意度モノラル化部 77 は、モノラル化処理の程度を強める。それによって、左右のチャンネルで常に一定以上の相関が確保できるようになる。

10

【0112】

音声認識機能を利用しようとする場合は、まず、先の図 7 の音声認識装置 7 と同様に、図 9 の起動命令部 792 により音声認識機能の起動信号が状態設定部 791 に送られ、状態設定部 791 は、切り替え部 78 を制御してスピーカ 9a および 9b への入力信号をステレオ信号から任意度モノラル化部 77 の出力に切り替え、音声認識の状態をスピーカ 9a および 9b にステレオ信号が入力された待機状態から、スピーカ 9a および 9b に任意度モノラル化信号が入力された動作状態に移行させる。モノラル度判定部 76 が常にオーディオ信号のモノラル度を監視しており、モノラル度が低いと判定された時にのみ、任意度モノラル化部 77 が任意度モノラル化処理を行う。モノラル度の低い一瞬の時間だけオーディオ信号のステレオ感が損なわれるが、十分なエコー打ち消し量が得られる。

20

【0113】

図 10 に、図 9 のモノラル度判定部 76 の詳細を示す。図 10 において、モノラル度判定部 76 は、加算器 81 と、減算器 82 と、レベル比較器 83 と、モノラル度計算部 84 とを含む。

【0114】

完全なモノラル信号の場合は、 $S_r = S_l$ であるから、加算器 81 の出力は $2S_r$ となり、減算器 82 の出力は "0" となり、 $\{ (\text{減算器 82 の出力レベル}) / (\text{加算器 81 の出力レベル}) \}$ を求めるレベル比較器 83 の出力も "0" になる。一方、完全なステレオ信号の場合、すなわち S_r と S_l とが完全に無相関である場合には、加算器 81 の出力は $S_r + S_l$ 、減算器 82 の出力は $S_r - S_l$ になる。 S_r と S_l とが完全に無相関であるから、 $S_r + S_l$ のレベルと、 $S_r - S_l$ のレベルとは等しく、レベル比較器 83 の出力は "1" になる。次に、モノラル度計算部 84 が $\{ 1 - (\text{レベル比較器出力}) \}$ を計算し、モノラル度判定部 76 は、完全なモノラル信号では "1" を、完全なステレオ信号では "0" を出力する。

30

【0115】

このように、モノラル度判定部 76 は、入力信号のモノラル度に応じて 1 ~ 0 の間の値を出力し、この値を監視することにより、入力信号のモノラル度を判定することができる。

【0116】

40

図 11 に、図 9 の任意度モノラル化部 77 の詳細を示す。図 11 において、任意度モノラル化部 77 は、処理強度決定部 91 と、減衰器 921 ~ 924 と、加算器 931 および 932 とを含む。

図 10 のモノラル度計算部 84 の出力が、図 11 の処理強度決定部 91 に入力され、処理強度決定部 91 は、この値に応じてモノラル化の処理強度を決定する。この処理強度に応じて、減衰器 921 ~ 924 の減衰量が制御される。

【0117】

図 12 は、図 11 の処理強度決定部 91 が行うモノラル化処理の強度、および図 11 の減衰器 921 ~ 924 を通じて実現される利得（減衰量）を示す図である。

図 12 (A) において、特性 101 は、図 11 の処理強度決定部 91 に入力されたモノラ

50

ル度と、処理強度決定部 9 1 から出力される処理強度との関係を示す。図 1 2 (B) において、特性 1 0 2 および特性 1 0 3 は、この処理強度決定部 9 1 から出力される処理強度により、減衰器 9 2 1 ~ 9 2 4 の利得がどのように制御されるかを示している。特性 1 0 2 が減衰器 9 2 1 および減衰器 9 2 4 の利得を示し、特性 1 0 3 が減衰器 9 2 2 および減衰器 9 2 3 の利得を示している。

【 0 1 1 8 】

本実施形態においては、処理強度決定部 9 1 は、入力信号のモノラル度が 1 . 0 ~ 0 . 5 の範囲では、特性 1 0 1 に示すように、減衰器 9 2 1 ~ 9 2 4 に対し、モノラル化処理強度として " 0 " を出力する。モノラル化処理強度が " 0 " の時、任意度モノラル化部 7 7 は、特性 1 0 2 と特性 1 0 3 とからかかるように、モノラル化処理を行わない。

10

【 0 1 1 9 】

処理強度決定部 9 1 は、入力信号のモノラル度が 0 . 5 以下になった場合にのみ " 0 " 以上のモノラル度処理強度を出力する。たとえばモノラル度 " 0 " の完全なステレオ信号が入力された場合は、処理強度決定部 9 1 は、減衰器 9 2 1 ~ 9 2 4 に対し、モノラル化処理強度として " 0 . 5 " を出力し、この時、任意度モノラル化部 7 7 からは、" 0 . 5 " のモノラル度をもつ信号が出力される。

【 0 1 2 0 】

図 1 2 に示す制御方式に従えば、モノラル化の処理強度が " 0 " の時には、右チャンネル信号が S_r 、左チャンネル信号が S_l になり、図 1 の A V 処理部 3 からのステレオ信号がそのままスピーカ 9 a および 9 b に入力される。処理強度が " 1 " の時には、両チャンネルとも $\{ (S_r + S_l) / 2 \}$ となり、完全にモノラル化された信号がスピーカ 9 a および 9 b に入力される。図 1 2 に示す特性では、処理強度の最大値が 0 . 5 に制限されている。このような範囲に制限する理由は、聴感上の自然性を確保しつつ、実用上十分なエコー打ち消し量を得られるようにするためである。

20

【 0 1 2 1 】

このようにモノラル化の程度を制限しても、音声認識の動作状態においては、短時間ではあるがオーディオ信号のステレオ感が損なわれる。そこで、音声認識機能の利用が終了すると同時に、終了命令部 7 9 3 から音声認識機能の終了信号が状態設定部 7 9 1 に送られ、状態設定部 7 9 1 は、切り替え部 7 8 を制御してスピーカ 9 a および 9 b への入力信号を任意度モノラル化部 7 7 の出力からステレオ信号へ切り替え、音声認識の状態を、スピーカ 9 a および 9 b に任意度モノラル化信号が入力された動作状態から、スピーカ 9 a および 9 b にステレオ信号が入力された待機状態に切り替える。

30

これにより、ある程度以上のステレオ感を確保しながら、常に十分なエコー打ち消し量を得ることができる。

【 0 1 2 2 】

以上のように、本実施形態によれば、音声認識機能の動作状態にあっても、普通のステレオ信号の場合はそのまま再生し、極端にモノラル度の低いステレオ信号だけに対してモノラル化処理を加えることにより、第 3 の実施形態よりも若干はエコー打ち消し効果は落ちるが、常に一定レベル以上のエコー打ち消し量を確保しつつ、ステレオ感の悪化をはるかに小さく抑えることができる。

40

【 0 1 2 3 】

(第 5 の実施形態)

図 1 3 は、本発明の第 5 の実施形態に係る A V 機器用音声認識装置の構成を示すブロック図である。図 1 3 の音声認識装置 7 は、図 1 の A V 機器に設けられた音声認識装置 7 と対応している。ただし、本実施形態では、A V 機器において、A V 処理部 3 から 4 チャンネル信号が出力され、スピーカユニット 9 に含まれる 4 つのスピーカ 9 a ~ 9 d を通じて 4 チャンネル音響が出力されるものとする。

【 0 1 2 4 】

図 1 3 において、音声認識装置 7 は、2 チャンネル化部 1 1 5 と、モノラル化部 1 1 6 と、1 個のエコーキャンセラ 1 1 3 と、音声認識部 1 1 4 と、起動命令部 1 1 9 2 と、終了

50

命令部 1 1 9 3 と、音声検出部 1 1 7 と、状態設定部 1 1 9 1 と、切り替え部 1 1 8 とを備えている。すなわち、図 1 3 の音声認識装置 7 は、図 7 の音声認識装置 7（第 3 の実施形態）において、2 つの選択肢の間で切り替えを行う切り替え部 5 6 を、3 つの選択肢の間で切り替えを行う切り替え部 1 1 8 に置き換えた上、2 チャンネル化部 1 1 5 と、音声検出部 1 1 7 とを追加したものである。音声検出部 1 1 7 は、図 5 の音声検出部 3 7（第 2 の実施形態を参照）と同じものである。スピーカ 9 a ~ 9 d へと入力される信号は、図 1 の A V 処理部 3 から出力される 4 チャンネル信号である。

【 0 1 2 5 】

スピーカ 9 a ~ 9 d へ向かう 4 チャンネル信号が分岐されて 2 チャンネル化部 1 1 5 に入力され、2 チャンネル化部 1 1 5 は、その 4 チャンネル信号を 2 チャンネル化する。2 チャンネル化部 1 1 5 の出力（以下、2 チャンネル化信号）は、モノラル化部 1 1 6 に入力され、モノラル化部 1 1 6 は、その 2 チャンネル化信号をモノラル化する。

10

【 0 1 2 6 】

マイクロホン 6 から出力される信号（マイクロホン出力）と、モノラル化部 1 1 6 から出力される信号（モノラル化信号）とがエコーキャンセラ 1 1 3 へと与えられ、エコーキャンセラ 1 1 3 は、そのマイクロホン出力からユーザの音声と対応する信号（以下、ユーザ音声）だけを抽出する。なお、エコーキャンセラ 1 1 3 の動作原理については、第 1 の実施形態で説明した。

【 0 1 2 7 】

起動命令部 1 1 9 2 は、音声認識動作の起動を命令する。終了命令部 1 1 9 3 は、音声認識動作の終了を命令する。状態設定部 1 1 9 1 は、起動命令部 1 1 9 2、終了命令部 1 1 9 3 からの命令を受け、音声認識部 1 1 4 の動作状態を設定（つまり音声認識動作を " O N " / " O F F " ）する。

20

音声検出部 1 1 7 へは、モノラル化部 1 1 6 の出力（モノラル化信号）と、エコーキャンセラ 1 1 3 の出力（ユーザ音声）とが与えられ、音声検出部 1 1 7 は、両出力のレベル比に基づいて、ユーザ音声を検出する。

【 0 1 2 8 】

モノラル化部 1 1 6 から出力される信号（モノラル化信号）と、2 チャンネル化部 1 1 5 から出力される信号（2 チャンネル化信号）と、図 1 の A V 処理部 3 からの 4 チャンネル信号とが切り替え部 1 1 8 へと与えられ、切り替え部 1 1 8 は、状態設定部 1 1 9 1 が音声認識動作を " O N " 状態に設定すると、スピーカ 9 a ~ 9 d への入力を、（ " O F F " 状態における）4 チャンネル信号から、2 チャンネル化信号に切り替える。さらに、この " O N " 状態において、音声検出部 1 1 7 がユーザ音声を検出すると、スピーカ 9 a ~ 9 d への入力を、（ " O N " 状態における）2 チャンネル化信号から、モノラル化信号に切り替える。また、切り替え部 1 1 8 は、状態設定部 1 1 9 1 が音声認識動作を " O F F " 状態に設定すると、スピーカ 9 a ~ 9 d への入力を、（ " O N " 状態における）2 チャンネル化信号またはモノラル化信号から、4 チャンネル信号に切り替える。

30

【 0 1 2 9 】

音声認識部 1 1 4 は、状態設定部 1 1 9 1 の設定に従い、音声認識を実行 / 終了する。すなわち、エコーキャンセラ 1 1 3 からのユーザ音声を認識して、その音声が表示コントロール信号を生成する。こうして生成されたコントロール信号は、図 1 のコントローラ 4 に伝達され、コントローラ 4 が受信部 2 や A V 処理部 3 を制御することによって、テレビジョン受像機において、受信チャンネルの切り換え等の処理が実行される。

40

【 0 1 3 0 】

図 1 3 の音声認識装置 7 のハードウェア的な構成は、図 3 と同様である。図 3 において、R O M 1 2 には、所定のプログラムが予め格納されている。このプログラムには、第 1 の実施形態で述べた（ a ） ~ （ c ）のアルゴリズムと、第 2 の実施形態で述べた（ d ）および（ e ）のアルゴリズム（ただしスピーカへの入力信号を切り替える際の選択肢の数は異なる）と、第 3 の実施形態で述べた（ f ）のアルゴリズムとに加え、さらに（ i ）4 チャンネル信号を 2 チャンネル化するためのアルゴリズムが記述されている。C P U 1 0 は、

50

R A M 1 1 を作業領域として利用しつつ、上記のプログラムに従って動作する。これによって、図 1 3 に示される各ブロックの機能が実現される。

【 0 1 3 1 】

なお、起動司令部 1 1 9 2 , 終了司令部 1 1 9 3 は、図 1 のコントローパネルを構成するボタンによって実現される。また、起動司令部 1 1 9 2 , 終了司令部 1 1 9 3 以外の各ブロックの機能をソフト的に実現する代わりに、各々専用のハード回路によって実現することもできる。

【 0 1 3 2 】

以上のように構成された A V 機器用音声認識装置 7 について、以下、その動作を述べる。4 チャンネルステレオでは、3 6 0 ° の音像定位を特徴とする音づくりがなされているため、チャンネル間の相関が極めて弱い。そのため、図 1 のスピーカユニット 9 に含まれる 4 つのスピーカ 9 a ~ 9 d を通じて 4 チャンネルステレオ信号を再生する場合、図 2 の音声認識装置 7 では、十分なエコー打ち消し量が得られず、その結果、正確な音声認識が行えないことが多い。そこで、図 5 の音声認識装置 7 や、図 7 の音声認識装置 7、図 9 の音声認識装置 7 のように、音声認識の動作状態においてのみ、あるいはユーザが音声を発した時のみ、スピーカ 9 a ~ 9 d からの音響をモノラル化して、必要なエコー打ち消し量を確保しようということになる。

10

【 0 1 3 3 】

ところが、4 チャンネルステレオ信号を一気にモノラル信号にまで変換してユーザに聴かせたのでは、ステレオ感（立体感）の喪失によるユーザの不満が極めて大きい。そこで、本実施形態では、音声認識の動作状態においては、4 チャンネル信号を 2 チャンネル化して 2 チャンネルステレオ音響をユーザに聴かせ、さらに、この動作状態においてユーザが音声を発した時のみ、モノラル化音響をユーザに聴かせる。これにより、動作状態においても、そこそこのステレオ感を保持しながら、十分なエコー打ち消し量を確保できる。

20

【 0 1 3 4 】

図 1 3 において、入力された 4 チャンネル信号から、2 チャンネル化部 1 1 5 で 2 チャンネル化信号が、モノラル化部 1 1 6 でモノラル化信号が生成される。

エコーキャンセラ 1 1 3 には、常にモノラル化信号が入力される。音声認識機能を利用しない待機状態では、4 チャンネル信号がスピーカ 9 a ~ 9 d に入力されている。

【 0 1 3 5 】

音声認識機能を利用しようとする場合は、まず、図 1 1 の起動司令部 1 1 9 2 により音声認識機能の起動信号が状態設定部 1 1 9 1 に送られ、状態設定部 1 1 9 1 は、切り替え部 1 1 8 を制御して、スピーカ 9 a ~ 9 d への入力信号を 4 チャンネル信号から 2 チャンネル化信号に切り替え、音声認識の状態をスピーカ 9 a ~ 9 d に 4 チャンネル信号が入力された待機状態から、スピーカ 9 a ~ 9 d に 2 チャンネル化信号が入力された動作状態に移行させる。4 チャンネルから 2 チャンネルへの変換は、右の前後チャンネルの信号を加算して右チャンネル信号とし、左の前後チャンネル信号を加算して左チャンネル信号とすることにより可能である。モノラル化は、上記 4 つのチャンネル信号を加算するか、あるいは左右 2 つの 2 チャンネル化信号を加算することにより行える。

30

【 0 1 3 6 】

動作状態においては、音声検出部 1 1 7 が、モノラル化信号とエコーキャンセラ出力信号のレベルを監視し、モノラル化信号から予想されるレベル以上にエコーキャンセラ出力信号のレベルが上昇した時に、ユーザが音声を発したと判断し、切り替え部 1 1 8 がスピーカ 9 a ~ 9 d の入力を 2 チャンネル化信号からモノラル化信号に切り替える。

40

【 0 1 3 7 】

以上のように、本実施形態によれば、音声認識機能を利用しない時は 4 チャンネル再生、音声認識の機能を働かせた時は 2 チャンネル再生、音声の入力時はモノラル再生というように、再生モードを段階的に切り替えることによって、待機状態においてもそこそこのステレオ感が確保でき、しかも、十分なエコー打ち消し量を得ることが可能となる。

【 0 1 3 8 】

50

(第6の実施形態)

図14は、本発明の第6の実施形態に係るAV機器用音声認識装置の構成を示すブロック図である。図14の音声認識装置7は、図1のAV機器に設けられた音声認識装置7と対応している。ただし、本実施形態では、AV機器において、AV処理部3から2チャンネル信号が出力され、スピーカユニット9に含まれる2つのスピーカ9aおよび9bを通じて2チャンネル音響が出力されるものとする。

【0139】

図14において、音声認識装置7は、モノラル化部125と、1個のエコーキャンセラ123と、音声認識部124と、起動司令部1282と、終了司令部1283と、状態設定部1281と、切り替え部127と、適応音発生部126とを備えている。すなわち、図14の音声認識装置7は、図7の音声認識装置7(第3の実施形態)に、適応音発生部126を追加したものである。スピーカ9aおよび9bへと入力される信号は、図1のAV処理部3から出力される2チャンネル信号である。

【0140】

適応音発生部126は、状態設定部1281の設定に関連して、モノラルの適応音を発生する。すなわち、状態設定部1281の設定によって音声認識動作が"OFF"状態から"ON"状態へと移行するのを受け、適応音発生部126は、モノラルの適応音を発生する。

【0141】

上記の適応音は、エコーキャンセラ123の適応動作を促進する作用を有する。すなわち、音声認識動作が"OFF"状態から"ON"状態へと移行するのに伴い、スピーカ9aおよび9bからの出力が2チャンネル音響からモノラル化音響へと切り替わるが、この切り替わり直後のスピーカ9aおよび9b出力のレベルが0(つまり無音響)または0に近い値であったとすると、エコーキャンセラ123では、2チャンネルに適応していたデジタルフィルタ123aのモノラルへの適応が進行しない。

【0142】

この時突然、スピーカ9aおよび9bから高レベルのモノラル化音響が出力されると、エコーキャンセラ123は、その音響を打ち消すことができない。その結果、音声認識部124に反響音が混入し、ユーザ音声を正確に認識できなくなる恐れがある。そこで、音声認識動作が"OFF"状態から"ON"状態へと移行する際、スピーカ9aおよび9bからモノラルの適応音響が出力されるようにして、デジタルフィルタ123を強制的にモノラルに適応させる。

適応音発生部126以外の構成要素の動作は、第3の実施形態と同様なので、説明を省略する。

【0143】

図14の音声認識装置7のハードウェア的な構成は、図3と同様である。図3において、ROM12には、所定のプログラムが予め格納されている。このプログラムには、第1の実施形態で述べた(a)~(c)のアルゴリズムと、第2の実施形態で述べた(e)のアルゴリズムと、第3の実施形態で述べた(f)のアルゴリズムとに加え、さらに(j)適応音を発生するためのアルゴリズム(あるいは適応音のサンプリングデータ)が記述されている。CPU10は、RAM11を作業領域として利用しつつ、上記のプログラムに従って動作する。これによって、図14に示される各ブロックの機能が実現される。

【0144】

なお、起動司令部1282、終了司令部1283は、図1のコントローパネルを構成するボタンによって実現される。また、起動司令部1282、終了司令部1283以外の各ブロックの機能をソフト的に実現する代わりに、各々専用のハード回路によって実現することもできる。

【0145】

以上のように構成されたAV機器用音声認識装置7について、以下、その動作を述べる。図14の音声認識装置7は、図7の音声認識装置7が持つ欠点、すなわち、音声認識の待

10

20

30

40

50

機状態から動作状態への移行直後において、エコーキャンセラ 5 4 のエコー打ち消し量が十分ではないという欠点を解決するものである。

【 0 1 4 6 】

図 7 の音声認識装置 7 では、音声認識の待機状態において、スピーカ 9 a および 9 b にはステレオ信号が、エコーキャンセラ 5 4 にはモノラル化信号が入力されているため、エコーキャンセラ 5 4 の適応は十分なものではなく、完全なエコー打ち消しが行えない。そこで、音声認識機能を利用しようとする時に、動作状態に移行してスピーカ 9 a および 9 b への入力信号をモノラル化信号に切り替えることによりエコーキャンセラ 5 4 を十分に適応させ、完全なエコー打ち消しを行わせるようにしていた。しかし、そのようにしても、スピーカ 9 a および 9 b から音が出ない限りは、エコーキャンセラ 5 4 の適応は進まない。従って、切り替え直後から長い無音区間が続き、ユーザが音声を入力中にスピーカ 9 a および 9 b から音が出始めた時には、スピーカ 9 a および 9 b からの反響音を十分に打ち消せないことになる。

10

【 0 1 4 7 】

そこで、図 1 4 の音声認識装置 7 では、待機状態から動作状態への移行の直後に、適応音発生部 1 2 6 からスピーカ 9 a および 9 b に、エコーキャンセラ 1 2 3 の適応を促進するためのモノラルの適応音を数秒間入力する構成としている。

適応音としては、「音声入力をお願いします」などの合成音声と考えられる。

【 0 1 4 8 】

以上のように、本実施形態によれば、待機状態から動作状態に移行した直後に、モノラルの適応音をスピーカ 9 a および 9 b から出力させることにより、移行直後であっても、十分なエコー打ち消し量を保証することが可能となる。

20

【 0 1 4 9 】

(第 7 の実施形態)

図 1 5 は、本発明の第 7 の実施形態に係る A V 機器用音声認識装置の構成を示すブロック図である。図 1 5 の音声認識装置 7 は、図 1 の A V 機器に設けられた音声認識装置 7 と対応している。ただし、本実施形態では、A V 機器において、A V 処理部 3 から 2 チャンネル信号が出力され、スピーカユニット 9 に含まれる 2 つのスピーカ 9 a および 9 b を通じて 2 チャンネル音響が出力されるものとする。

【 0 1 5 0 】

図 1 5 において、音声認識装置 7 は、モノラル化部 1 3 5 と、1 個のエコーキャンセラ 1 3 3 と、音声認識部 1 3 4 と、起動命令部 1 3 8 2 と、終了命令部 1 3 8 3 と、状態設定部 1 3 8 1 と、切り替え部 1 3 6 と、打ち消し監視部 1 3 7 1 と、音声検出部 1 3 7 2 と、減衰部 1 3 7 3 とを備えている。すなわち、図 1 5 の音声認識装置 7 は、図 7 の音声認識装置 7 (第 3 の実施形態) に、打ち消し監視部 1 3 7 1 と、音声検出部 1 3 7 2 と、減衰部 1 3 7 3 とを追加したものである。スピーカ 9 a および 9 b へと入力される信号は、図 1 の A V 処理部 3 から出力される 2 チャンネル信号である。

30

【 0 1 5 1 】

打ち消し監視部 1 3 7 1 へは、モノラル化部 1 3 5 の出力 (モノラル化信号) と、エコーキャンセラ 1 3 3 の出力 (ユーザ音声) とが与えられ、各出力のレベル変動を監視することにより、エコーキャンセラ 1 3 3 において反響音が十分打ち消されているか (つまり、デジタルフィルタ 1 3 3 a のモノラルへの適応が十分進んでいるか) を判定する。すなわち、モノラル化信号のレベルが急激に立ち上がった時、ユーザ音声のレベルも急に立ち上がれば、反響音は十分には打ち消されていないことになり、逆に、ほとんど立ち上がらなければ、反響音は十分に打ち消されているといえる。

40

【 0 1 5 2 】

減衰部 1 3 7 3 は、打ち消し監視部 1 3 7 1 の監視結果、および状態設定部 1 3 8 1 の設定に関連して、図 1 の A V 処理部 3 から入力される 2 チャンネル信号を減衰させる。すなわち、エコーキャンセラ 1 3 3 において反響音が十分には打ち消されていない時に、状態設定部 1 3 8 1 の設定によって音声認識動作が " O F F " 状態から " O N " 状態へと移行

50

するのを受け、減衰部 1 3 7 3 は 2 チャンネル信号を一時的に減衰させる。

【 0 1 5 3 】

上記のようにして 2 チャンネル信号を減衰させれば、音声認識部 1 3 4 への反響音の混入を防止することができる。すなわち、音声認識動作が " O F F " 状態から " O N " 状態へと移行するのに伴い、スピーカ 9 a および 9 b からの出力が 2 チャンネル音響からモノラル化音響へと切り替わるが、この切り替わり直後のスピーカ出力のレベルが 0 (つまり無音響) または 0 に近い値であったとすると、エコーキャンセラ 1 3 3 では、2 チャンネルに適用していたデジタルフィルタ 1 3 3 a のモノラルへの適用が進行しない。

【 0 1 5 4 】

この時突然、スピーカ 9 a および 9 b から高レベルのモノラル化音響が出力されると、エコーキャンセラ 1 3 3 は、その音響を打ち消すことができない。そこで、反響音が十分には打ち消されていない場合には、音声認識動作が " O F F " 状態から " O N " 状態へと移行する際に、スピーカ 9 a および 9 b から出力されるモノラル化信号のレベルを低下させることによって、音声認識部 1 3 4 への反響音の混入を防止する。

打ち消し監視部 1 3 7 1、音声検出部 1 3 7 2 および減衰部 1 3 7 3 以外の構成要素の動作は、第 3 の実施形態と同様なので、説明を省略する。

【 0 1 5 5 】

図 1 5 の音声認識装置 7 のハードウェア的な構成は、図 3 と同様である。図 3 において、ROM 1 2 には、所定のプログラムが予め格納されている。このプログラムには、第 1 の実施形態で述べた (a) ~ (c) のアルゴリズムと、第 2 の実施形態で述べた (e) のアルゴリズムと、第 3 の実施形態で述べた (f) のアルゴリズムとに加え、さらに (k) 反響音が十分打ち消されているかを監視するためのアルゴリズム、および (l) スピーカへの 2 チャンネル信号を減衰させるためのアルゴリズムが記述されている。CPU 1 0 は、RAM 1 1 を作業領域として利用しつつ、上記のプログラムに従って動作する。これによって、図 1 5 に示される各ブロックの機能が実現される。

【 0 1 5 6 】

なお、起動命令部 1 3 8 2、終了命令部 1 3 8 3 は、図 1 のコントローパネルを構成するボタンによって実現される。また、起動命令部 1 3 8 2、終了命令部 1 3 8 3 以外の各ブロックの機能をソフト的に実現する代わりに、各々専用のハード回路によって実現することもできる。

【 0 1 5 7 】

以上のように構成された A V 機器用音声認識装置 7 について、以下、その動作を述べる。図 1 5 の音声認識装置 7 は、先の図 1 4 の音声認識装置 7 と同様、図 7 の音声認識装置 7 が持つ欠点、すなわち、待機状態から動作状態への移行直後において、エコーキャンセラ 5 4 のエコー打ち消し量が十分ではないという欠点を解決するものである。

【 0 1 5 8 】

図 1 5 の音声認識装置 7 には、前述のように、モノラル化部 1 3 5 の出力信号レベルと、エコーキャンセラ 1 3 3 の出力信号レベルとを監視して、反響音が十分打ち消されているかを判定する打ち消し監視部 1 3 7 1 と、モノラル化部 1 3 5 の出力信号レベルと、エコーキャンセラ 1 3 3 の出力信号レベルとを監視して、ユーザが音声を発したか否かを判断する音声検出部 1 3 7 2 と、スピーカ 9 a および 9 b への入力信号を減衰させる減衰部 1 3 7 3 とが設けられている。待機状態から動作状態に移行した直後は、エコーキャンセラ 1 3 3 の適用が完全ではなく、当然、エコーキャンセラ 1 3 3 におけるエコー打ち消し効果も悪い。もし、動作状態への移行後、スピーカ 9 a および 9 b への入力信号において無音区間が続き、エコーキャンセラ 1 3 3 が適用できないうちに、ユーザが音声を発し、しかも、ユーザが音声を発している最中にスピーカ 9 a および 9 b から音が出始めた時、エコーキャンセラ 1 3 3 は、スピーカ 9 a および 9 b からの反響音を十分に打ち消すことができず、音声認識部 1 3 4 への音声にスピーカ 9 a および 9 b からの反響音が混入することになる。

【 0 1 5 9 】

そこで、本実施形態では、音声検出部 1 3 7 2 と減衰部 1 3 7 3 とを設け、モノラル化信号において無音区間が続いており、エコーキャンセラ 1 3 3 によって反響音が十分打ち消されてないと打ち消し監視部 1 3 7 1 が判定し、しかも音声検出部 1 3 7 2 がユーザ音声を検出した時、減衰部 1 3 7 3 によりスピーカ 9 a および 9 b への入力信号を減衰させ、それによって、ユーザ音声への反響音の混入を少なくしている。ユーザが音声を発しない状態でスピーカ 9 a および 9 b へのモノラル化信号が無音から有音となった場合には、減衰部 1 3 7 3 の減衰量を " 0 " とし、スピーカ 9 a および 9 b から出力されるモノラル化音響を適応音として、エコーキャンセラ 1 3 3 の適応を促進する。また、エコーキャンセラ 1 3 3 の適応が進み、残存エコーが小さくなった場合には、音声検出時においても減衰量は " 0 " に制御される。

10

【 0 1 6 0 】

以上のように、本実施形態によれば、待機状態（音声認識動作が " O F F " の状態）から動作状態（ " O N " の状態）に移行した直後のエコー打ち消し量が十分でない状態においてユーザが音声を発した時、これを検知してスピーカ 9 a および 9 b への入力信号に適切な減衰を挿入して、スピーカ 9 a および 9 b からの音響のレベルを下げ、それによって反響音の混入を防止するので、エコー打ち消し量が十分でない状態における音声認識性能が高まる。

【 0 1 6 1 】

（第 8 の実施形態）

図 1 6 は、本発明の第 8 の実施形態に係る A V 機器用音声認識装置の構成を示すブロック図である。図 1 6 の音声認識装置 7 は、図 1 の A V 機器に設けられた音声認識装置 7 と対応している。ただし、本実施形態では、A V 機器において、A V 処理部 3 から 2 チャンネル信号が出力され、スピーカユニット 9 に含まれる 2 つのスピーカ 9 a および 9 b を通じて 2 チャンネル音響が出力されるものとする。

20

【 0 1 6 2 】

図 1 6 において、音声認識装置 7 は、モノラル化部 1 4 5 と、1 個のエコーキャンセラ 1 4 3 と、音声認識部 1 4 4 と、起動命令部 1 4 8 2 と、終了命令部 1 4 8 3 と、状態設定部 1 4 8 1 と、切り替え部 1 4 6 と、適応制御部 1 4 7 とを備えている。すなわち、図 1 6 の音声認識装置 7 は、図 7 の音声認識装置 7 （第 3 の実施形態）に、適応制御部 1 4 7 を追加したものである。スピーカ 9 a および 9 b へと入力される信号は、図 1 の A V 処理部 3 から出力される 2 チャンネル信号である。

30

【 0 1 6 3 】

適応制御部 1 4 7 は、状態設定部 1 4 8 1 の設定に関連して、エコーキャンセラ 1 4 3 内の適応デジタルフィルタ 1 4 3 a の適応速度を制御する。すなわち、デジタルフィルタ 1 4 3 a は、入力信号への適応速度が可変であり、適応制御部 1 4 7 は、モノラル用の速い適応速度と、2 チャンネル用の遅い適応速度とを予め記憶している。そして、状態設定部 1 4 8 1 の設定によって音声認識動作が " O F F " 状態から " O N " 状態へと移行する（それに伴い、スピーカ出力が 2 チャンネル音響からモノラル音響に切り替わる）のを受け、適応制御部 1 4 7 は、デジタルフィルタ 1 4 3 a の適応速度を、遅い適応速度から速い適応速度に変更する。また、音声認識動作が " O N " 状態から " O F F " 状態へと移行するのを受け、デジタルフィルタ 1 4 3 a の適応速度を、速い適応速度から遅い適応速度に変更する。

40

適応制御部 1 4 7 以外の構成要素の動作は、第 3 の実施形態と同様なので、説明を省略する。

【 0 1 6 4 】

図 1 6 の音声認識装置 7 のハードウェア的な構成は、図 3 と同様である。図 3 において、R O M 1 2 には、所定のプログラムが予め格納されている。このプログラムには、第 1 の実施形態で述べた（ a ）～（ c ）のアルゴリズムと、第 2 の実施形態で述べた（ e ）のアルゴリズムと、第 3 の実施形態で述べた（ f ）のアルゴリズムとに加え、さらに（ m ）エコーキャンセラの適応速度を制御するためのアルゴリズムが記述されている。C P U 1 0

50

は、RAM 11 を作業領域として利用しつつ、上記のプログラムに従って動作する。これによって、図 16 に示される各ブロックの機能が実現される。

【0165】

なお、起動命令部 1482、終了命令部 1483 は、図 1 のコントローパネルを構成するボタンによって実現される。また、起動命令部 1482、終了命令部 1483 以外の各ブロックの機能をソフト的に実現する代わりに、各々専用のハード回路によって実現することもできる。

【0166】

以上のように構成された AV 機器用音声認識装置 7 について、以下、その動作を述べる。よく知られているように、エコーキャンセラ 143 の適応は、その出力が "0" になる方向に、推定インパルスレスポンスを逐次修正していくものである。スピーカ 9a および 9b とマイクロホン 6 との間の系（反響路）のインパルスレスポンスは、家具、人、窓、カーテンなどの影響を受けて刻々と変化するため、適応動作なくしては満足な打ち消し量が得られない。ただし、騒音のように、インパルスレスポンスをいかに修正しても消去できない信号がエコーキャンセラ 143 への入力信号に含まれている場合には、推定インパルスレスポンスにおいて誤差が生じ、この誤差がエコー打ち消し量を悪化させる。

【0167】

推定インパルスレスポンスの逐次修正において、1 回当たりの修正量を変えることにより、適応速度を制御することができる。1 回当たりの修正量が大きいと適応速度が速くなり、修正量が小さいと適応速度が遅くなる。適応速度を速めると、系の変動には強い（つまり反響路のインパルスレスポンス変動に素早く追従できる）が、雑音には弱くなる（つまり雑音によって適応動作が不安定となりやすい）。逆に、適応速度を遅くすると、系の変動には弱くなるが、雑音には強くなる。従って、実際の装置では、系の変動への追従性と耐雑音性との両方をそこそこ満足するような適応速度が選ばれていた。

【0168】

図 7 の音声認識装置 7 では、音声認識の動作状態においては、スピーカ 9a および 9b からモノラル化された音響が出力され、そのモノラル化音響の反響音をモノラル化信号で打ち消すため、比較的速い適応速度でも良好な動作が可能である。ところが、待機状態においては、ステレオ音響の反響音を、モノラル化信号で打ち消すため、動作状態と同じ適応速度では、推定インパルスレスポンスに含まれる誤差が極めて大きくなる。原理的に打ち消し不可能な信号成分を打ち消そうとエコーキャンセラ 143 が適応動作を重ねるため、せっかく推定したインパルスレスポンスを破壊してしまうことになる。このように、図 7 の音声認識装置 7 は、待機状態における適応性能が悪いため、待機状態から動作状態への移行直後のエコー打ち消し量が極めて少ないという欠点を持っていた。

【0169】

そこで、図 16 の音声認識装置 7 では、エコーキャンセラ 143 の適応速度を制御する適応制御部 147 を設けることにより、待機状態から動作状態に切り替わった直後においても、十分なエコー打ち消し量が得られるようにした。

すなわち、適応制御部 147 は、エコーキャンセラ 143 に対し、スピーカ 9a および 9b にステレオ信号が入力される待機状態と、モノラル化信号が入力される動作状態とで、互いに異なる適応速度を設定する。

具体的には、適応制御部 147 は、待機状態では、適応速度を遅くして、インパルスレスポンスの推定精度を確保する。一方、動作状態では、適応速度を速めることにより、待機状態から動作状態への移行直後においても、十分なエコー打ち消し効果が得られるようにする。

【0170】

以上のように、本実施形態によれば、エコーキャンセラ 143（内の適応デジタルフィルタ 143a）の適応速度を、音声認識部 144 が動作状態に設定されている時には速い速度に、待機状態に設定されている時には遅い速度に制御するので、モノラル、マルチチャンネルそれぞれに好適なエコー打ち消しを行うことができる。

すなわち、スピーカ 9 a , 9 b から出力されるのがマルチチャンネル音響の場合は、適応デジタルフィルタ 1 4 3 a から見ると雑音であるステレオ成分が多いので、遅い適応速度とすることによって耐雑音性を高め、一方、モノラル化音響の場合は、ステレオ成分がないので、速い適応速度とすることによって、反響路のインパルスレスポンスの変動への追従性を高めることができる。

【 0 1 7 1 】

また、エコーキャンセラ 1 4 3 の適応速度を、上記のように音声認識動作の状態に応じて変化させることにより、待機状態から動作状態への移行直後においても、優れたエコー打ち消し効果が実現できる。

【 0 1 7 2 】

(第 9 の実施形態)

図 1 7 は、本発明の第 9 の実施形態に係る A V 機器用音声認識装置の構成を示すブロック図である。図 1 7 の音声認識装置 7 は、図 1 の A V 機器に設けられた音声認識装置 7 と対応している。ただし、本実施形態では、A V 機器において、A V 処理部 3 から 2 チャンネル信号が出力され、スピーカユニット 9 に含まれる 2 つのスピーカ 9 a および 9 b を通じて 2 チャンネル音響が出力されるものとする。

【 0 1 7 3 】

図 1 7 において、音声認識装置 7 は、モノラル化部 1 5 5 と、1 個のエコーキャンセラ 1 5 3 と、音声認識部 1 5 4 と、起動司令部 1 5 8 2 と、終了司令部 1 5 8 3 と、状態設定部 1 5 8 1 と、切り替え部 1 5 6 と、適応制御部 1 5 7 とを備えている。すなわち、図 1 7 の音声認識装置 7 は、図 1 6 の音声認識装置 7 (第 8 の実施形態) と同様の構成を有する。図 1 6 の音声認識装置 7 との相違は、次の点である。すなわち、図 1 の A V 処理部 3 からの音響信号は、2 チャンネル (ステレオ) の場合と、モノラルの場合とがあり、図 1 の A V 処理部 3 から適応制御部 1 5 7 へ、モノラル / ステレオ識別信号がさらに与えられる。スピーカ 9 a および 9 b へと入力される信号は、図 1 の A V 処理部 3 から出力される 2 チャンネルまたはモノラル信号である。

【 0 1 7 4 】

適応制御部 1 5 7 は、状態設定部 1 5 8 1 の設定と、モノラル / ステレオ識別信号とに関連して、エコーキャンセラ 1 5 3 内の適応デジタルフィルタ 1 5 3 a の適応速度を制御する。すなわち、デジタルフィルタ 1 5 3 a は、入力信号への適応速度が可変であり、適応制御部 1 5 7 は、モノラル用の速い適応速度と、2 チャンネル用の遅い適応速度とを予め記憶している。状態設定部 1 5 8 1 の設定によって音声認識動作が " O F F " 状態から " O N " 状態へと移行する (それに伴い、スピーカ出力が 2 チャンネル音響からモノラル音響に切り替わる) のを受け、適応制御部 1 5 7 は、デジタルフィルタ 1 5 3 a の適応速度を、遅い適応速度から速い適応速度に変更する。また、音声認識動作が " O N " 状態から " O F F " 状態へと移行するのを受け、デジタルフィルタ 1 5 3 a の適応速度を、速い適応速度から遅い適応速度に変更する。

【 0 1 7 5 】

ただし、適応制御部 1 5 7 は、モノラル / ステレオ識別信号がステレオを示している場合のみ、上記のような適応速度の変更を行い、モノラルを示す場合には、状態設定部 1 5 8 1 の設定に関わらず、デジタルフィルタ 1 5 3 a の適応速度を速い適応速度とする。適応制御部 1 5 7 以外の構成要素の動作は、第 8 の実施形態と同様なので、説明を省略する。

【 0 1 7 6 】

図 1 7 の音声認識装置 7 のハードウェア的な構成は、図 3 と同様である。図 3 において、R O M 1 2 に格納されているプログラムは、第 8 の実施形態と同様である。ただし、上記 (m) のアルゴリズム、すなわちエコーキャンセラの適応速度を制御するためのアルゴリズムについては、音声認識動作の " O N " / " O F F " 状態だけでなく、モノラル / ステレオ識別信号をも参照して制御を行うように変更が加えられている。C P U 1 0 は、R A M 1 1 を作業領域として利用しつつ、上記のプログラムに従って動作する。これによって

10

20

30

40

50

、図１７に示される各ブロックの機能が実現される。

【０１７７】

なお、起動命令部１５８２，終了命令部１５８３は、図１のコントローパネルを構成するボタンによって実現される。また、起動命令部１５８２，終了命令部１５８３以外の各ブロックの機能をソフト的に実現する代わりに、各々専用のハード回路によって実現することもできる。

【０１７８】

以上のように構成されたＡＶ機器用音声認識装置７について、以下、その動作を述べる。一般のＴＶ放送においては、ステレオ番組とモノラル番組の２つの番組があり、ステレオ番組かモノラル番組かを識別する識別信号が、映像／音響信号とともに放送されている。受信側では、この識別信号により、現在の番組がステレオ番組かモノラル番組かを知ることができる。

10

先の図１６の音声認識装置７では、現在受信しているのがステレオ番組かモノラル番組かを問わず、モノラル化部１５５で処理した信号がスピーカ９ａおよび９ｂに入力されていない待機状態においては、エコーキャンセラ１５３の適応速度を落としていたが、待機状態であっても、適応速度を落とすたくないのは当然である。

【０１７９】

適応速度を落とした状態では、エコーキャンセラ１５３が系の変動に追従できていない可能性があり、そのような時に動作状態に移行すると、十分なエコー打ち消し量が得られない。これに対して、待機状態であっても適応速度を落とさなければ、エコーキャンセラ１５３が常に系の変動に追従できているので、いつ動作状態に移行しても、十分なエコー打ち消し量が確保できるからである。

20

【０１８０】

放送そのものがモノラル番組であれば、モノラル化部１５５でモノラル化していない待機状態においても、適応速度を速めることが可能である。そこで、図１７の音声認識装置７では、適応制御部１５７は、まず識別信号をチェックし、その結果、現在受信しているのがステレオ番組の場合、待機状態では、エコーキャンセラ１５３の適応速度を遅くするが、モノラル番組の場合、待機状態であっても動作状態と同様、適応速度を速いままに保つ。

【０１８１】

30

以上のように、本実施形態によれば、ステレオ／モノラル識別信号に基づいて、現在受信している番組の音声ステレオかモノラルかを判定して、モノラルの場合は、音声認識動作が待機状態であっても、エコーキャンセラ１５３の適応速度を遅くしないので、反響路のインパルスレスポンスの変動への追従性が低下することがなく、その結果、待機状態において優れたエコー打ち消し効果が実現でき、動作状態へ移行した直後における音声認識性能が高まる。

【０１８２】

（第１０の実施形態）

図１８は、本発明の第１０の実施形態に係るＡＶ機器用音声認識装置の構成を示すブロック図である。図１８の音声認識装置７は、図１のＡＶ機器に設けられた音声認識装置７と対応している。ただし、本実施形態では、ＡＶ機器において、ＡＶ処理部３から２チャンネル信号が出力され、スピーカユニット９に含まれる２つのスピーカ９ａおよび９ｂを通じて２チャンネル音響が出力されるものとする。

40

【０１８３】

図１８において、音声認識装置７は、モノラル化部１６５と、１個のエコーキャンセラ１６３と、音声認識部１６４と、起動命令部１６８２と、終了命令部１６８３と、状態設定部１６８１と、切り替え部１６６と、モノラル度判定部１６７１と、適応制御部１６７２とを備えている。すなわち、図１８の音声認識装置７は、図１６の音声認識装置７（第８の実施形態）に、モノラル度判定部１６７１を追加したものである。モノラル度判定部１６７１は、図９のモノラル度判定部７６（第４の実施形態を参照）と同じものである。ス

50

ピーカ 9 a および 9 b へと入力される信号は、図 1 の A V 処理部 3 から出力される 2 チャンネル信号である。

【 0 1 8 4 】

モノラル度判定部 1 6 7 1 へは、上記の 2 チャンネル信号が分岐入力され、モノラル度判定部 1 6 7 1 は、その 2 チャンネル信号のモノラル度を判定する。適応制御部 1 6 7 2 は、モノラル判定部の判定結果に関連して、エコーキャンセラ 1 6 3 内の適応デジタルフィルタ 1 6 3 a の適応速度を制御する。

【 0 1 8 5 】

すなわち、適応制御部 1 6 7 2 は、2 チャンネル信号のモノラル度に応じて、デジタルフィルタ 1 6 3 a の適応速度を変化させる。好ましくは、モノラル度が高いほど、適応速度を速くする。そのために、適応制御部 1 6 7 2 は、適応速度を速める処理をどの強度で行うべきかをモノラル度をもとに決定するための関数（処理強度決定特性；図 1 9 に参照番号 1 0 4 で示される）を記憶している。

モノラル度判定部 1 6 7 1 および適応制御部 1 6 7 2 以外の構成要素の動作は、第 8 の実施形態と同様なので、説明を省略する。

【 0 1 8 6 】

図 1 8 の音声認識装置 7 のハードウェア的な構成は、図 3 と同様である。図 3 において、ROM 1 2 には、所定のプログラムが予め格納されている。このプログラムには、第 1 の実施形態で述べた（a）～（c）のアルゴリズムと、第 2 の実施形態で述べた（e）のアルゴリズムと、第 3 の実施形態で述べた（f）のアルゴリズムと、第 4 の実施形態で述べた（g）のアルゴリズムと、第 8 の実施形態で述べた（m）のアルゴリズムとが記述されている。

【 0 1 8 7 】

ただし、上記（m）のアルゴリズム、すなわちエコーキャンセラの適応速度を制御するためのアルゴリズムについては、音声認識動作の“ON”/“OFF”状態に基づいて制御を行う（第 8 の実施形態）のではなく、スピーカへの 2 チャンネル信号のモノラル度に基づいて制御を行うように変更が加えられている。CPU 1 0 は、RAM 1 1 を作業領域として利用しつつ、上記のプログラムに従って動作する。これによって、図 1 8 に示される各ブロックの機能が実現される。

【 0 1 8 8 】

なお、起動司令部 1 6 8 2 , 終了司令部 1 6 8 3 は、図 1 のコントローパネルを構成するボタンによって実現される。また、起動司令部 1 6 8 2 , 終了司令部 1 6 8 3 以外の各ブロックの機能をソフト的に実現する代わりに、各々専用のハード回路によって実現することもできる。

【 0 1 8 9 】

以上のように構成された A V 機器用音声認識装置 7 について、以下、その動作を述べる。図 1 8 の音声認識装置 7 は、先の図 7 の音声認識装置 7 において、モノラル度の低い信号が入力された時に、エコーキャンセラ 1 6 3 の適応精度が劣化するという欠点を解決するものである。前述したように、エコーキャンセラ 1 6 3 の適応は、その出力が“0”になる方向に、推定インパルスレスポンスを逐次修正していくものであり、騒音のように、インパルスレスポンスをいかに修正しても消去できない信号がエコーキャンセラ 1 6 3 への入力信号に含まれている場合には、推定インパルスレスポンスにおいて誤差が生じ、この誤差がエコー打ち消し量を悪化させる。

【 0 1 9 0 】

これと同じことが、ステレオ信号の反響音を、モノラル化信号で打ち消す場合に起きる。すなわち、ステレオ信号の反響音をモノラル化信号で打ち消す場合、原理的に、インパルスレスポンスをいかに修正しても打ち消せない成分が残る。

この打ち消せない成分（ステレオ成分）が多い場合、すなわち、モノラル度の低いステレオ信号の場合には、原理的に打ち消せない信号を打ち消そうとエコーキャンセラ 1 6 3 が適応動作を重ねるため、せっかく推定したインパルスレスポンスを大きく破壊することに

10

20

30

40

50

なる。

【0191】

そこで、図18の音声認識装置7では、AV処理部3からのステレオ信号を分析して、原理的にエコー打ち消しが精度よく行え適応動作に適した信号であるかどうかを判断し、適応動作に適すると判断された時に、エコーキャンセラ163に適応動作を行わせる。

【0192】

図18の音声認識装置7では、適応に適した信号か否かは、信号のモノラル度で判断している。前述したように、モノラル度の高い信号ほどエコー打ち消し効果が高く、インパルスレスポンスの推定がうまく行える。そこで、まず、モノラル度判定部1671がステレオ信号のモノラル度を求める。次に、適応制御部1672がこのモノラル度に応じて、エコーキャンセラ163の適応速度を制御する。

10

【0193】

図19は、図18の適応制御部1672が行う適応速度制御処理の特性を示す図である。図19において、特性191は、図18のスピーカ9aおよび9bへ向かうステレオ信号のモノラル度と、エコーキャンセラ163の適応速度との関係を示している。図19からわかるように、適応制御部1672は、ステレオ信号のモノラル度が高く適応に適すると判断した場合には、適応速度を上げて常に最良の推定インパルスレスポンスを得られるようにする。一方、モノラル度が低く適応に適さないと判断した場合には、適応速度を下げて推定インパルスレスポンスの破壊を防止する。

20

【0194】

以上のように、本実施形態によれば、2チャンネル信号（ステレオ信号）のモノラル度に基づいて、適応デジタルフィルタ163aの適応速度を制御するので、様々なモノラル度を持った2チャンネル信号に好適なエコー打ち消しを行うことができる。

すなわち、モノラル度が低い場合、適応速度を遅くして、耐雑音性を高める。

一方、モノラル度が高い場合、適応デジタルフィルタ163aから見ると雑音であるステレオ成分が少ないので、耐雑音性はあまり必要でない。よって、適応速度を速くすることによって、反響路のインパルスレスポンスの変動への追従性を高めることができる。その結果、特にモノラル度が高い場合に、優れたエコー打ち消し効果の実現でき、動作状態へ移行した直後における音声認識性能が高まる。

30

【0195】

（第11の実施形態）

図20は、本発明の第11の実施形態に係るAV機器用音声認識装置の構成を示すブロック図である。図20の音声認識装置7は、図1のAV機器に設けられた音声認識装置7と対応している。ただし、本実施形態では、AV機器において、AV処理部3から2チャンネル信号が出力され、スピーカユニット9に含まれる2つのスピーカ9aおよび9bを通じて2チャンネル音響が出力されるものとする。

【0196】

図20において、音声認識装置7は、モノラル化部175と、1個のエコーキャンセラ173と、音声認識部174と、起動命令部1782と、終了命令部1783と、状態設定部1781と、切り替え部176と、不揮発メモリ177とを備えている。すなわち、図20の音声認識装置7は、図7の音声認識装置7（第3の実施形態）に、不揮発メモリ177を追加したものである。スピーカ9aおよび9bへと入力される信号は、図1のAV処理部3から出力される2チャンネル信号である。

40

【0197】

不揮発メモリ177へは、図1のコントロールパネル5からの電源"ON"/"OFF"信号が与えられ、不揮発メモリ177は、電源"OFF"時、エコーキャンセラ173が保持している推定インパルスレスポンスを取得し、それを記憶する。そして、電源"ON"時、記憶している推定インパルスレスポンスを、エコーキャンセラ173（内の適応デジタルフィルタ173a）に与える。

エコーキャンセラ173は、反響音を打ち消す動作を開始する際に、この不揮発メモリ1

50

77から与えられた推定インパルスレスポンスを初期値として用いる。すなわち、適応デジタルフィルタ173aは、不揮発メモリ177から与えられた値を初期値としてインパルスレスポンスの推定を開始する。

【0198】

エコーキャンセラ173は、電源"ON"時に用いる初期値の違いを除けば、図7のエコーキャンセラ54(第3の実施形態)と同様の動作を行う。なお、エコーキャンセラ54の場合、反響音を打ち消す動作を開始する際、"0"を初期値として用いるので、電源"ON"の直後から、デジタルフィルタ54aの適応が進むまでの期間、反響音が十分に打ち消されない問題があった。

不揮発メモリ177およびエコーキャンセラ173以外の構成要素の動作は、第3の実施形態と同様なので、説明を省略する。

10

【0199】

図20の音声認識装置7のハードウェア的な構成は、図3において、さらに不揮発メモリ177を追加したものである。ROM12には、所定のプログラムが予め格納されている。このプログラムには、第1の実施形態で述べた(a)~(c)のアルゴリズムと、第2の実施形態で述べた(e)のアルゴリズムと、第3の実施形態で述べた(f)のアルゴリズムとに加え、さらに(n)電源"OFF"時にエコーキャンセラ173が保持している推定インパルスレスポンスを不揮発メモリ177に書き込み、かつ電源"ON"時、その推定インパルスレスポンスをエコーキャンセラ173に与える手順が記述されている。CPU10は、RAM11を作業領域として利用しつつ、上記のプログラムに従って動作する。これによって、図20に示される各ブロックの機能が実現される。

20

【0200】

なお、起動司令部1782、終了司令部1783は、図1のコントローパネルを構成するボタンによって実現される。また、起動司令部1782、終了司令部1783以外の各ブロックの機能をソフト的に実現する代わりに、各々専用のハード回路によって実現することもできる。

【0201】

以上のように構成されたAV機器用音声認識装置7について、以下、その動作を述べる。スピーカ9aおよび9bからマイクロホン6までの反響路のインパルスレスポンスは、壁、天井、床、家具、人、窓、カーテンなどでの音響の反射状態によって決まる。同じAV機器であっても、設置環境により千差万別のインパルスレスポンスが得られる。しかも、AV機器の移動、家具の移動、人の出入り、窓の開閉などにより時々刻々と変化する。固定したインパルスレスポンスでは、十分なエコー打ち消し効果が得られない。このため、図7の音声認識装置7のエコーキャンセラ173は、逐次適応を行っており、常に最新のインパルスレスポンスを推定している。しかし、インパルスレスポンスの初期値を"0"とした適応方法では、電源"ON"の直後に十分なエコー打ち消し量が得られないという欠点があった。

30

【0202】

人とか窓といった細かい変化を除けば、AV機器の設置位置や部屋形状などで決まる大まかなインパルスレスポンスは、部屋の家具の配置換えなどをしなければ、今日昨日で大きく変わるものではない。昨日の電源"OFF"時の推定インパルスレスポンスを、今日の電源"ON"時に使用しても、そこそこのエコー打ち消し量が得られる場合が多い。

40

【0203】

そこで、図20の音声認識装置7では、不揮発メモリ177を設け、電源"OFF"時にエコーキャンセラ173が保持していた推定インパルスレスポンスを不揮発メモリ177に記憶させ、電源"ON"時には、この不揮発メモリ177が記憶している推定インパルスレスポンスを初期値として、エコーキャンセラ173をスタートさせる。

【0204】

以上のように、本実施形態によれば、電源"OFF"時の推定インパルスレスポンスを記憶しておき、電源"ON"時、それを初期値としてインパルスレスポンスの推定を開始す

50

るので、" 0 " を初期値とする場合と比べ、電源 " O N " 直後の推定誤差が小さくなり、その結果、音声認識性能が高まる。

【 0 2 0 5 】

(第 1 2 の実施形態)

図 2 1 は、本発明の第 1 2 の実施形態に係る A V 機器用音声認識装置の構成を示すブロック図である。図 2 1 の音声認識装置 7 は、図 1 の A V 機器に設けられた音声認識装置 7 と対応している。ただし、本実施形態では、A V 機器において、A V 処理部 3 から 2 チャンネル信号が出力され、スピーカユニット 9 に含まれる 2 つのスピーカ 9 a および 9 b を通じて 2 チャンネル音響が出力されるものとする。

【 0 2 0 6 】

図 2 1 において、音声認識装置 7 は、モノラル化部 1 8 5 と、1 個のエコーキャンセラ 1 8 3 と、音声認識部 1 8 4 と、音声検出部 1 8 7 と、起動司令部としてのボタンスイッチ 1 8 8 2 と、終了司令部としての時限スイッチ 1 8 8 3 と、状態設定部 1 8 8 1 と、切り替え部 1 8 6 とを備えている。すなわち、図 2 1 の音声認識装置 7 は、図 7 の音声認識装置 7 (第 3 の実施形態) において、音声検出部 1 8 7 を追加し、さらに、起動司令部 5 8 1 を特にボタンスイッチ 1 8 8 2 とし、かつ終了司令部 5 8 2 を特に時限スイッチ 1 8 8 3 としたものである。音声検出部 1 8 7 は、図 5 の音声検出部 3 7 と同じものである (第 2 の実施形態を参照) 。スピーカ 9 a および 9 b へと入力される信号は、図 1 の A V 処理部 3 から出力される 2 チャンネル信号である。

【 0 2 0 7 】

ボタンスイッチ 1 8 8 2 が押されると、ボタンスイッチ 1 8 8 2 から状態設定部 1 8 8 1 へ、音声認識動作の起動を命令する信号が送られる。音声検出部 1 8 7 は、ユーザ音声の有無を検出して、検出結果を時限スイッチ 1 8 8 3 に通知する。時限スイッチ 1 8 8 3 は、ユーザの音声の有無の状態から無の状態へ移行した瞬間を捉えて計時処理を開始する。そして、計時開始から予め決められた時間が経過すると、状態設定部 1 8 8 1 へ、音声認識動作の終了を命令する信号を送る。

【 0 2 0 8 】

状態設定部 1 8 8 1 は、ボタンスイッチ 1 8 8 2 , 時限スイッチ 1 8 8 3 からの命令信号を受け、音声認識部 1 8 4 の動作状態を設定 (つまり音声認識動作を " O N " / " O F F ") する。

音声検出部 1 8 7 、ボタンスイッチ 1 8 8 2 、時限スイッチ 1 8 8 3 および状態設定部 1 8 8 1 以外の構成要素の動作は、第 3 の実施形態と同様なので、説明を省略する。

【 0 2 0 9 】

図 2 1 の音声認識装置 7 のハードウェア的な構成は、図 3 と同様である。図 3 において、R O M 1 2 には、所定のプログラムが予め格納されている。このプログラムには、第 1 の実施形態で述べた (a) ~ (c) のアルゴリズムと、第 2 の実施形態で述べた (e) のアルゴリズムと、第 3 の実施形態で述べた (f) のアルゴリズムとに加え、さらに (o) 計時を行い、かつ計時開始から予め決められた時間が経過すると終了命令信号を送信する手順が記述されている。C P U 1 0 は、R A M 1 1 を作業領域として利用しつつ、上記のプログラムに従って動作する。これによって、図 2 1 に示される各ブロックの機能が実現される。

【 0 2 1 0 】

なお、ボタンスイッチ 1 8 8 2 は、図 1 のコントローパネルを構成するいずれかのボタンによって実現される。また、ボタンスイッチ 1 8 8 2 以外の各ブロックの機能をソフト的に実現する代わりに、各々専用のハード回路によって実現することもできる。

【 0 2 1 1 】

以上のように構成された A V 機器用音声認識装置 7 について、以下、その動作を述べる。本実施形態では、図 7 の音声認識装置 7 において、起動司令部 5 8 1 および終了司令部 5 8 2 の具体例を示している。ユーザは、音声認識機能を利用しようとする場合、まず、図 7 の起動司令部 5 8 1 に相当するボタンスイッチ 1 8 8 2 を押す。すると、状態設定部 1

10

20

30

40

50

８８１に対して待機状態（音声認識動作が“OFF”の状態）から動作状態（“ON”状態）への切り替えの指示がなされ、かつ、時限スイッチ１８８３に対して時間計測開始の指示がなされる。

【０２１２】

動作状態においては、音声検出部１８７がユーザ音声が入力されたか否かをチェックし、音声を検出された時に、時限スイッチ１８８３は、計測時間をリセットする（つまり計測時間を０に戻す）。音声を検出されない状態が続き、時限スイッチ１８８３の計測時間が定められた値を超えた時、時限スイッチ１８８３は、状態設定部１８８１に動作状態から待機状態への切り替えを指示する。

【０２１３】

以上のように、本実施形態によれば、音声認識機能の終了を自動的に行えるようになる。

【０２１４】

（第１３の実施形態）

図２２は、本発明の第１３の実施形態に係るＡＶ機器用音声認識装置の構成を示すブロック図である。図２２の音声認識装置７は、図１のＡＶ機器に設けられた音声認識装置７と対応している。ただし、本実施形態では、ＡＶ機器において、ＡＶ処理部３から２チャンネル信号が出力され、スピーカユニット９に含まれる２つのスピーカ９ａおよび９ｂを通じて２チャンネル音響が出力されるものとする。

【０２１５】

図２２において、音声認識装置７は、モノラル化部１９５と、１個のエコーキャンセラ１９３と、音声認識部１９４と、音声検出部１９７と、起動命令部としての音声スイッチ１９８２と、終了命令部としての時限スイッチ１９８３と、状態設定部１９８１と、切り替え部１９６とを備えている。すなわち、図２２の音声認識装置７は、図７の音声認識装置７（第３の実施形態）において、音声検出部１９７を追加し、さらに、起動命令部５８１を特に音声スイッチ１９８２とし、かつ終了命令部５８２を特に時限スイッチ１９８３としたものである。音声検出部１９７は、図５の音声検出部３７と同じものである（第２の実施形態を参照）。スピーカ９ａおよび９ｂへと入力される信号は、図１のＡＶ処理部３から出力される２チャンネル信号である。

【０２１６】

音声検出部１９７は、ユーザ音声の有無を検出して、検出結果を音声スイッチ１９８２および時限スイッチ１９８３に通知する。音声スイッチ１９８２は、ユーザ音声が無から有に移行した瞬間を捉えて、状態設定部１９８１へ音声認識動作の起動を命令する信号を送る。時限スイッチ１９８３は、ユーザの音声の有無の状態から無の状態へ移行した瞬間を捉えて計時処理を開始する。そして、計時開始から予め決められた時間が経過すると、状態設定部１９８１へ、音声認識動作の終了を命令する信号を送る。

【０２１７】

状態設定部１９８１は、音声スイッチ１９８２，時限スイッチ１９８３からの命令信号を受け、音声認識部１９４の動作状態を設定（つまり音声認識動作を“ON”/“OFF”）する。

音声検出部１９７、音声スイッチ１９８２、時限スイッチ１９８３および状態設定部１９８１以外の構成要素の動作は、第３の実施形態と同様なので、説明を省略する。

【０２１８】

図２２の音声認識装置７のハードウェア的な構成は、図３と同様である。図３において、ＲＯＭ１２には、所定のプログラムが予め格納されている。このプログラムには、第１の実施形態で述べた（ａ）～（ｃ）のアルゴリズムと、第２の実施形態で述べた（ｅ）のアルゴリズムと、第３の実施形態で述べた（ｆ）のアルゴリズムと、第１２の実施形態で述べた（ｏ）の手順に加え、さらに（ｐ）音声を検出されると起動命令信号を送信する手順が記述されている。ＣＰＵ１０は、ＲＡＭ１１を作業領域として利用しつつ、上記のプログラムに従って動作する。これによって、図２１に示される各ブロックの機能が実現される。

10

20

30

40

50

【 0 2 1 9 】

なお、各ブロックの機能をソフト的に実現する代わりに、各々専用のハード回路によって実現することもできる。

【 0 2 2 0 】

以上のように構成された A V 機器用音声認識装置 7 について、以下、その動作を述べる。
図 2 2 の音声認識装置 7 では、待機状態にあっても音声検出部 1 9 7 がユーザの音声の検出している。ユーザが音声認識機能を利用しようとする場合、まず、比較的大きな声を発する。音声検出部 1 9 7 がこの音声を検出し、検出結果を音声スイッチ 1 9 8 2 に送る。検出結果があらかじめ設定したレベル以上の音声を検出したことを示す時、音声スイッチ 1 9 8 2 は、音声認識の開始命令を状態設定部 1 9 8 1 に送り、状態設定部 1 9 8 1 に待機状態から動作状態への切り替えを指示する。

10

【 0 2 2 1 】

音声検出部 1 9 7 による検出結果は、時限スイッチ 1 9 8 3 へも送られ、応じて、時限スイッチ 1 9 8 3 は、時間計測を開始する。動作状態においては、音声検出部 1 9 7 がユーザの音声が入力されたか否かをチェックし、音声を検出された時に、時限スイッチ 1 9 8 3 は、計測時間をリセットする（つまり計測時間を 0 に戻す）。音声を検出されない状態が続く、時限スイッチ 1 9 8 3 の計測時間が定められた値を超えた時、時限スイッチ 1 9 8 3 は、状態設定部 1 9 8 1 に動作状態から待機状態への切り替えを指示する。

【 0 2 2 2 】

上記の音声スイッチ 1 9 8 2 が " O N " となる音声レベルは、時限スイッチ 1 9 8 3 がリセットされる音声レベルよりもかなり高く設定される。エコーキャンセラ 1 9 3 の打ち消し効果がよくない待機状態で発生する比較的大きなレベルの消し残しの反響音がユーザ音声として誤検出され、それに伴って動作モードに移行することがないようにするためである。

20

【 0 2 2 3 】

以上のように、本実施形態によれば、音声認識機能の開始と終了を自動的に行えるようになる。

【図面の簡単な説明】

【図 1】本発明が用いられる A V 機器の構成の一例を示すブロック図である。

【図 2】本発明の第 1 の実施形態に係る A V 機器用音声認識装置の構成を示すブロック図である。

30

【図 3】図 1 の音声認識装置 7 のハードウェア的な構成を示すブロック図である。

【図 4】図 2 の音声認識装置 7 において、各構成要素に入力される、または各構成要素から出力される信号の時間波形を示す図である。

【図 5】本発明の第 2 の実施形態に係る A V 機器用音声認識装置の構成を示すブロック図である。

【図 6】図 5 の音声認識装置 7 において、各構成要素に入力される、または各構成要素から出力される信号の時間波形を示す図である。

【図 7】本発明の第 3 の実施形態に係る A V 機器用音声認識装置の構成を示すブロック図である。

40

【図 8】図 7 の音声認識装置 7 において、各構成要素に入力される、または各構成要素から出力される信号の時間波形を示す図である。

【図 9】本発明の第 4 の実施形態に係る A V 機器用音声認識装置の構成を示すブロック図である。

【図 1 0】図 9 のモノラル度判定部 7 6 の詳細を示すブロック図である。

【図 1 1】図 9 の任意度モノラル化部 7 7 の詳細を示すブロック図である。

【図 1 2】図 1 1 の処理強度決定部 9 1 が行うモノラル化処理の強度、および図 1 1 の減衰器 9 2 1 ~ 9 2 4 を通じて実現される利得（減衰量）を示す図である。

【図 1 3】本発明の第 5 の実施形態に係る A V 機器用音声認識装置の構成を示すブロック図である。

50

【図 1 4】本発明の第 6 の実施形態に係る A V 機器用音声認識装置の構成を示すブロック図である。

【図 1 5】本発明の第 7 の実施形態に係る A V 機器用音声認識装置の構成を示すブロック図である。

【図 1 6】本発明の第 8 の実施形態に係る A V 機器用音声認識装置の構成を示すブロック図である。

【図 1 7】本発明の第 9 の実施形態に係る A V 機器用音声認識装置の構成を示すブロック図である。

【図 1 8】本発明の第 1 0 の実施形態に係る A V 機器用音声認識装置の構成を示すブロック図である。

10

【図 1 9】図 1 8 の適応制御部 1 6 7 2 が行う適応速度制御処理の特性を示す図である。

【図 2 0】本発明の第 1 1 の実施形態に係る A V 機器用音声認識装置の構成を示すブロック図である。

【図 2 1】本発明の第 1 2 の実施形態に係る A V 機器用音声認識装置の構成を示すブロック図である。

【図 2 2】本発明の第 1 3 の実施形態に係る A V 機器用音声認識装置の構成を示すブロック図である。

【図 2 3】従来の A V 機器用音声認識装置の構成を示すブロック図である。

【図 2 4】図 2 3 の音声認識装置において、各構成要素に入力される、または各構成要素から出力される信号の時間波形を示す図である。

20

【図 2 5】別の従来の A V 機器用音声認識装置の構成を示すブロック図である。

【符号の説明】

6 ... マイクロホン

7 ... A V 機器用音声認識装置

9 a ~ 9 c ... スピーカ

1 3 , 3 3 他 ... モノラル化部

1 4 , 3 4 他 ... エコーキャンセラ

1 4 a , 3 4 a 他 ... 適応デジタルフィルタ

1 4 b , 3 4 b 他 ... 減算回路

1 5 , 3 5 他 ... 音声認識部

30

3 6 , 5 6 他 ... 切り替え部

3 7 , 1 1 7 他 ... 音声検出部

5 7 , 7 9 1 他 ... 状態設定部

5 8 1 , 7 9 2 他 ... 起動命令部

5 8 2 , 7 9 3 他 ... 終了命令部

7 5 ... 完全モノラル化部

7 6 , 1 6 7 1 ... モノラル度判定部

7 7 ... 任意度モノラル化部

1 1 5 ... 2 チャンネル化部

1 2 6 ... 適応音発生部

40

1 3 7 3 ... 減衰部

1 3 7 1 ... 打ち消し監視部

1 4 7 , 1 5 7 , 1 6 7 2 ... 適応制御部

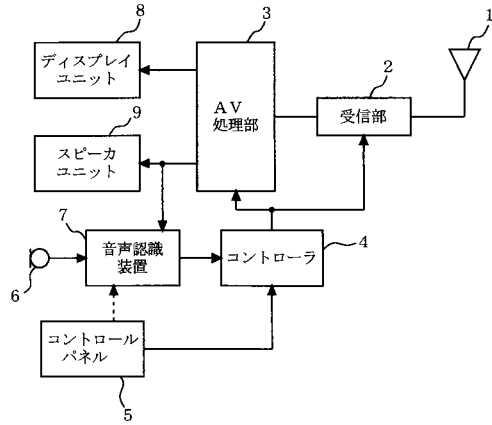
1 7 7 ... 不揮発メモリ

1 8 8 2 ... ボタンスイッチ

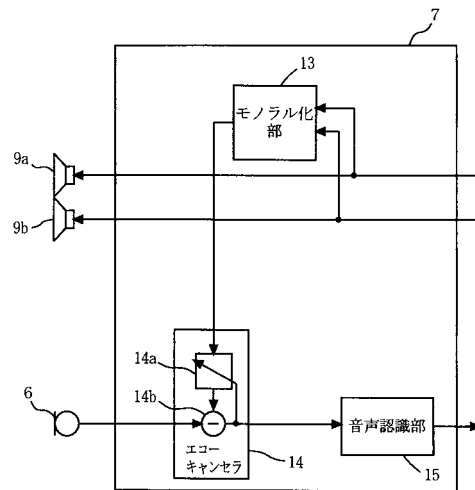
1 8 8 3 , 1 9 8 3 ... 時限スイッチ

1 9 8 2 ... 音声スイッチ

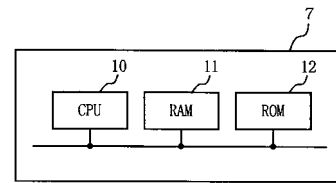
【図 1】



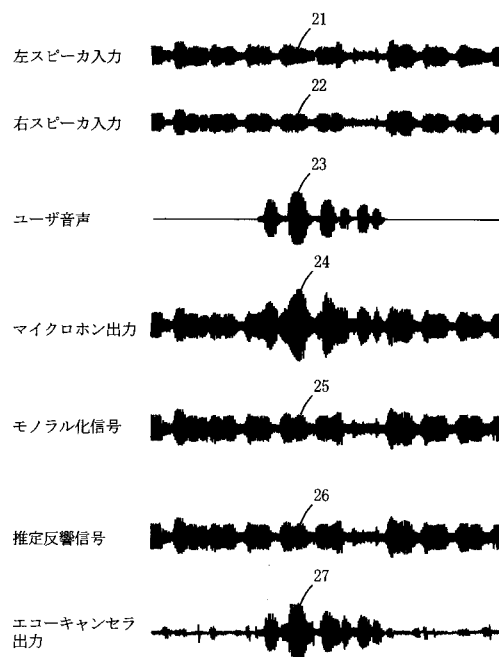
【図 2】



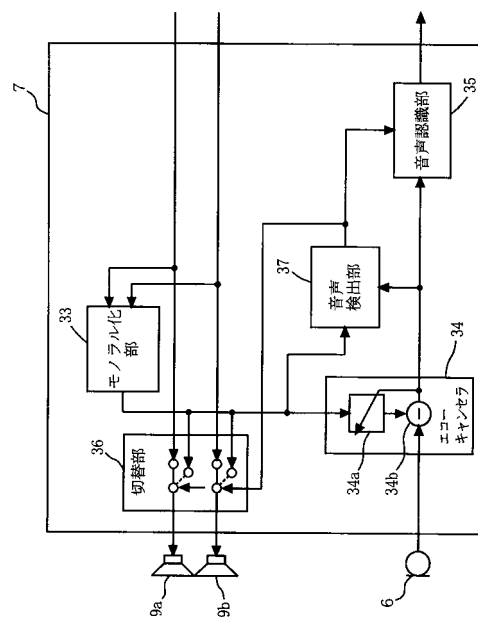
【図 3】



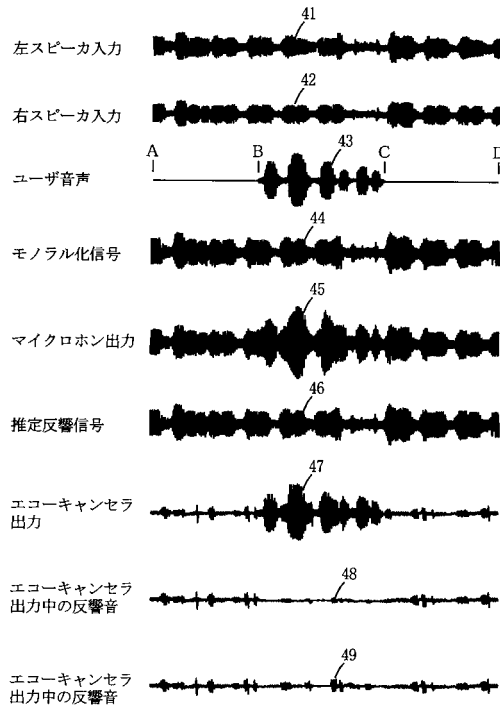
【図 4】



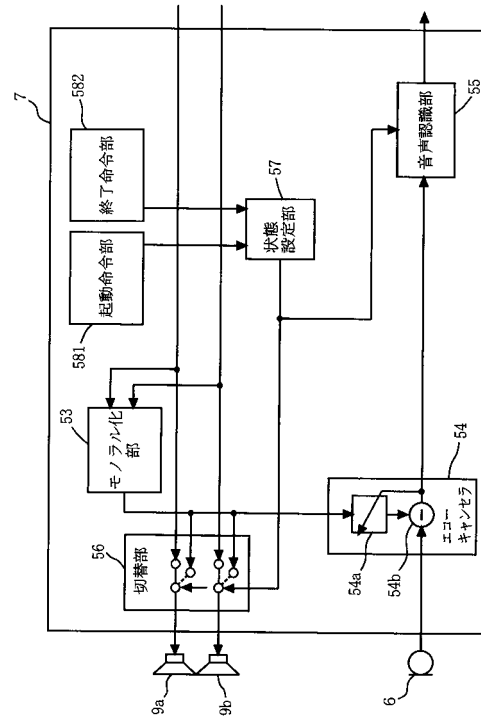
【図 5】



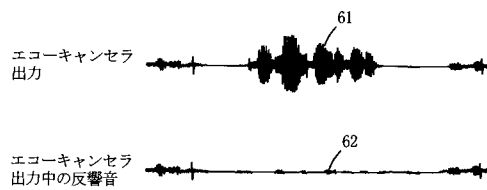
【図 6】



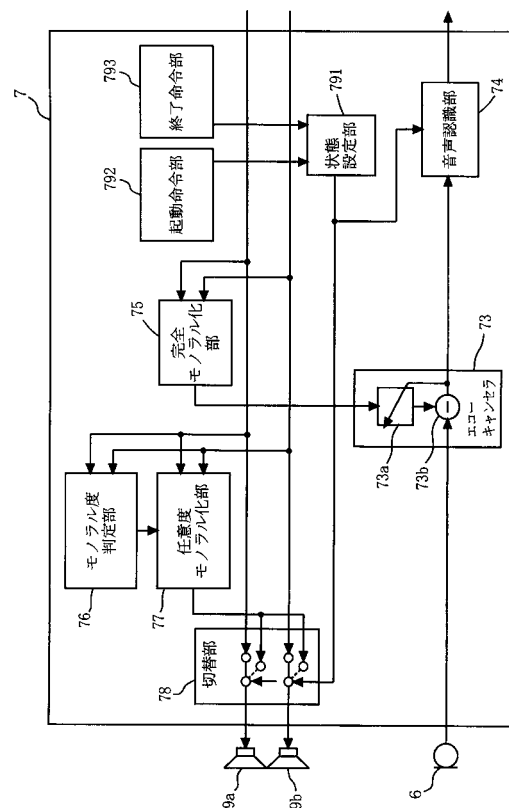
【図 7】



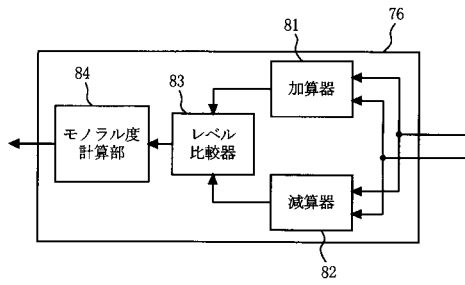
【図 8】



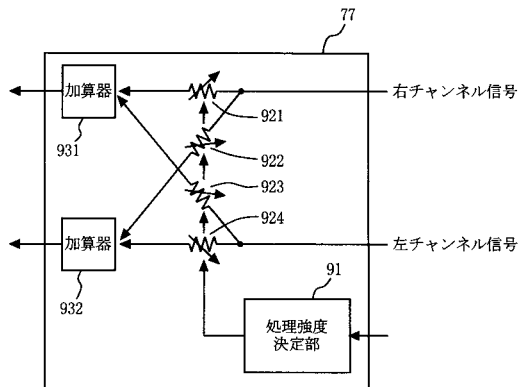
【図 9】



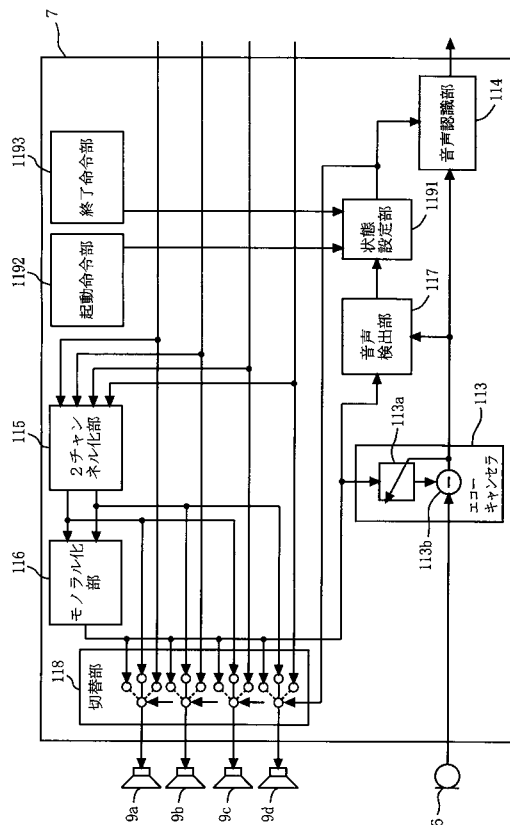
【 図 1 0 】



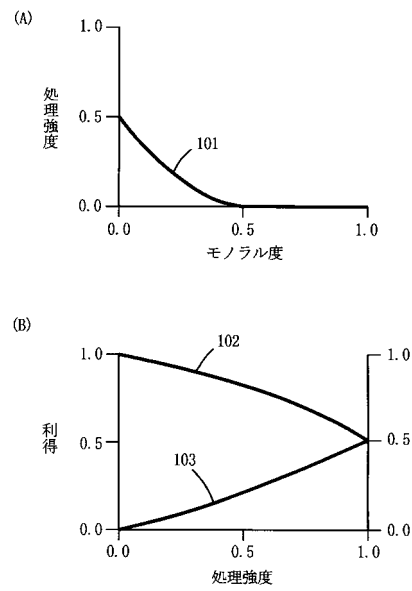
【 図 1 1 】



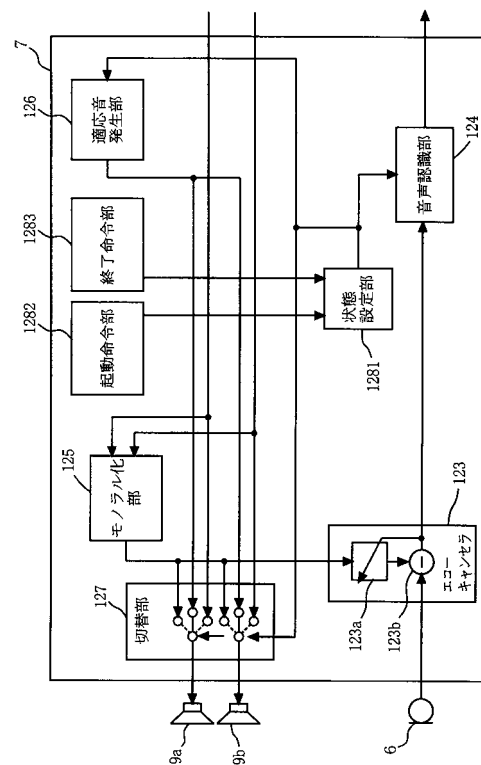
【 図 1 3 】



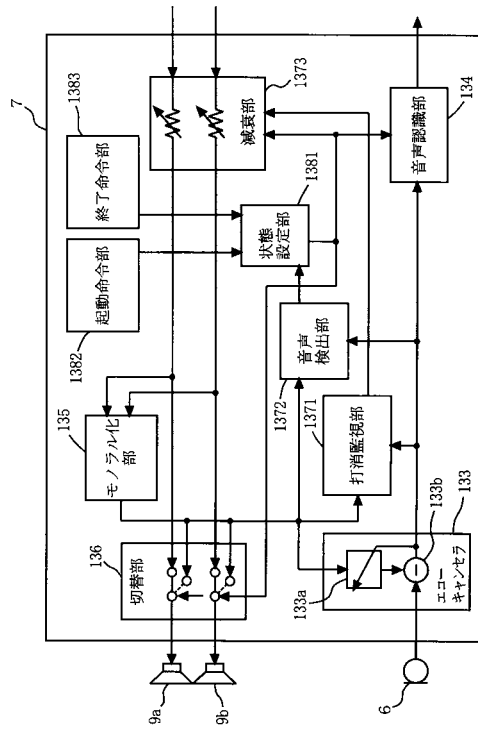
【圖 12】



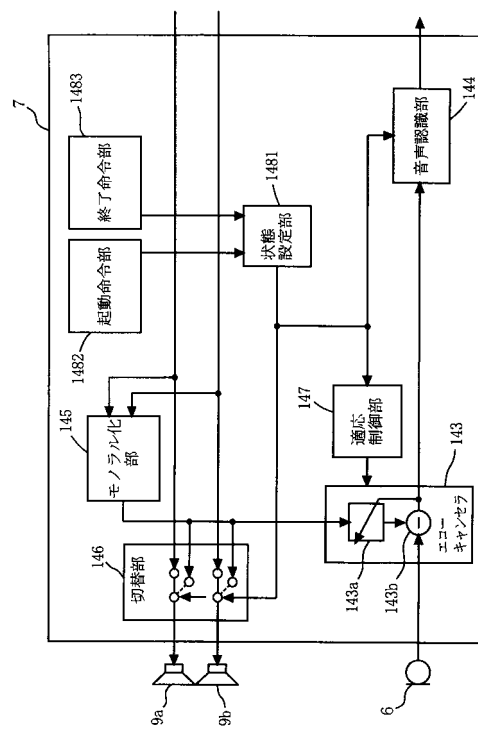
【 図 1 4 】



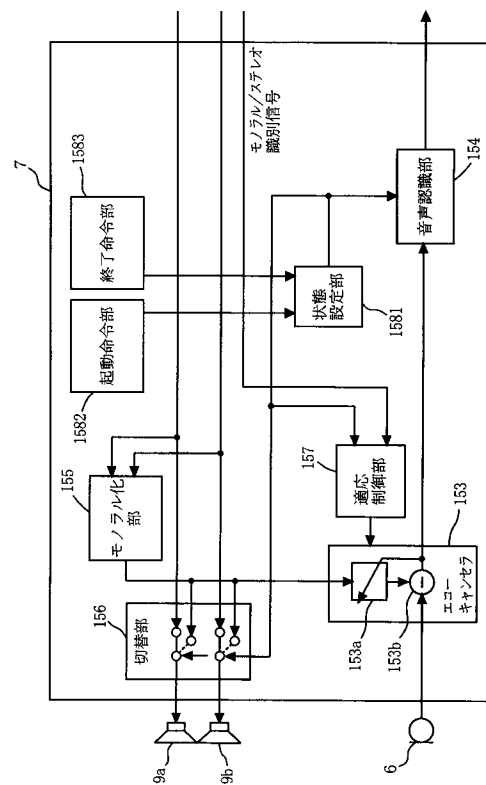
【 図 1 5 】



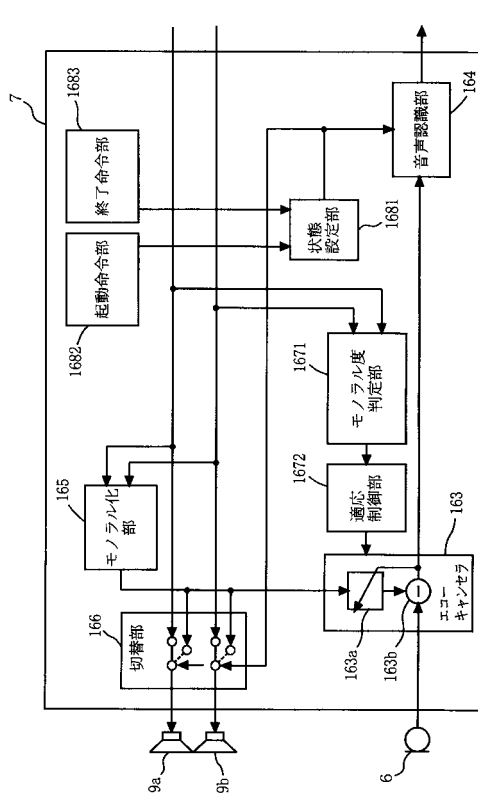
【 図 1 6 】



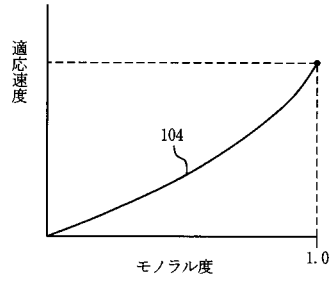
【圖 17】



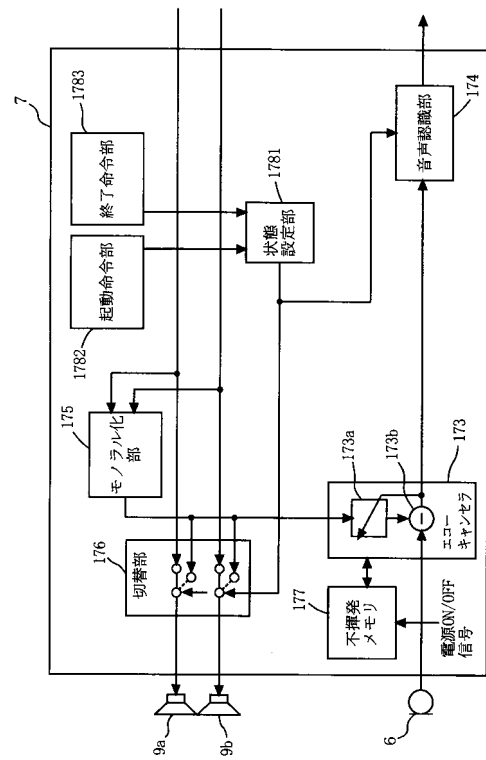
【 図 1 8 】



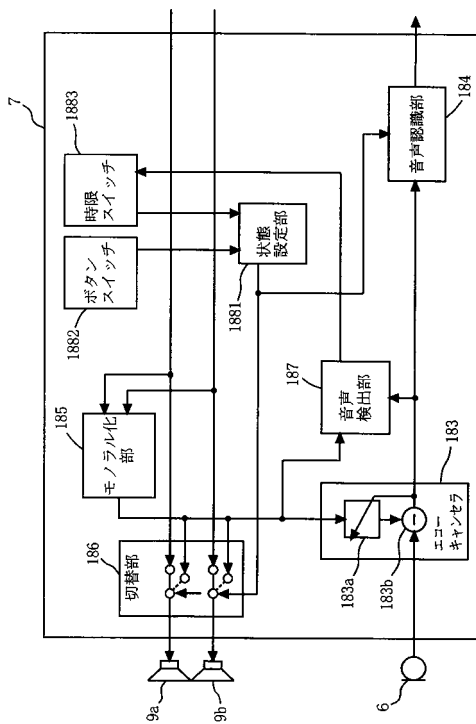
【図 19】



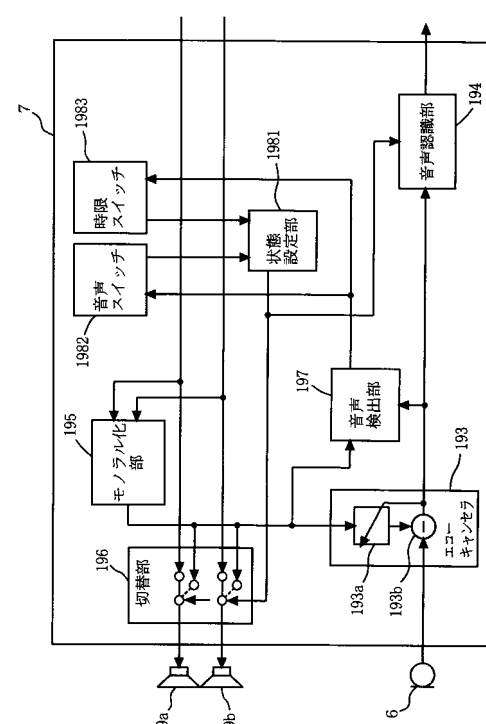
【図 20】



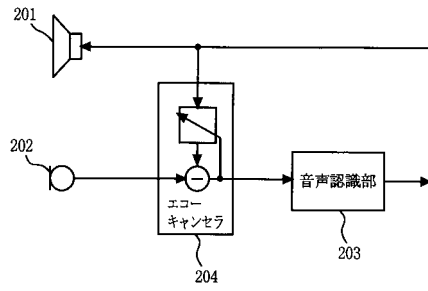
【図 21】



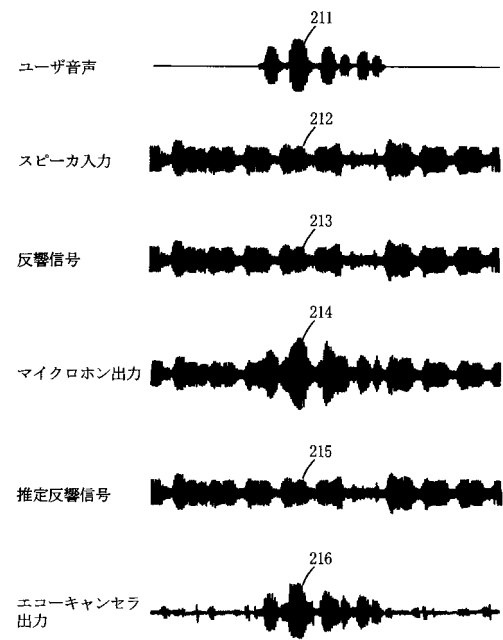
【図 22】



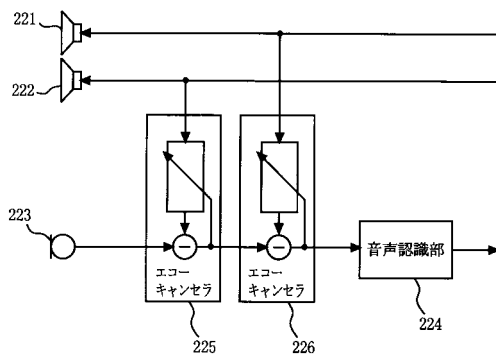
【図 23】



【図 24】



【図 25】



フロントページの続き

(72)発明者 小田 幹夫
大阪府門真市大字門真 1 0 0 6 番地 松下電器産業株式会社内

審査官 山下 剛史

(56)参考文献 特開平 0 2 - 2 4 4 0 9 9 (J P , A)
特開平 0 7 - 0 6 4 5 9 8 (J P , A)

(58)調査した分野(Int.Cl. , D B 名)

G10L 15/00-15/28

H04B 3/23

H04R 3/02, 3/12

H04S 1/00