

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5695447号
(P5695447)

(45) 発行日 平成27年4月8日(2015.4.8)

(24) 登録日 平成27年2月13日(2015.2.13)

(51) Int.Cl.		F I			
G 1 0 L	15/00	(2013.01)	G 1 0 L	15/00	2 0 0 G
G 1 0 L	15/28	(2013.01)	G 1 0 L	15/28	2 3 0 Z
H 0 4 N	7/173	(2011.01)	H 0 4 N	7/173	6 3 0

請求項の数 6 (全 15 頁)

(21) 出願番号	特願2011-44430 (P2011-44430)	(73) 特許権者	000003078
(22) 出願日	平成23年3月1日(2011.3.1)		株式会社東芝
(65) 公開番号	特開2012-181374 (P2012-181374A)		東京都港区芝浦一丁目1番1号
(43) 公開日	平成24年9月20日(2012.9.20)	(74) 代理人	100149803
審査請求日	平成25年10月25日(2013.10.25)		弁理士 藤原 康高
		(72) 発明者	大内 一成
			東京都港区芝浦一丁目1番1号 株式会社東芝内
		(72) 発明者	河村 聡典
			東京都港区芝浦一丁目1番1号 株式会社東芝内
		(72) 発明者	酒井 優
			東京都港区芝浦一丁目1番1号 株式会社東芝内

最終頁に続く

(54) 【発明の名称】 テレビジョン装置及び遠隔操作装置

(57) 【特許請求の範囲】

【請求項1】

音声を入力する音声入力部と、
 音声認識を開始する指示の入力をユーザから受ける指示入力部と、
 前記指示の入力を受けた後のユーザの発話に対して音声認識を行って音声認識結果を得る音声認識部と、
 前記音声認識結果に対応付けられた操作コマンドを実行する制御部と、
放送番組を録画して録画コンテンツを得る録画部と、
録画コンテンツを再生する再生部と、
 を有し、
 前記制御部は、前記指示の入力を受けた時の出力音量の設定値が閾値以上である場合、前記音声認識部が音声認識を行う間の出力音量の設定値を一時的に前記閾値以下の値に変更し、
前記放送番組の視聴中に前記指示の入力を受けた場合、前記録画部は前記放送番組の録画を開始し、
前記音声認識部による音声認識処理が終わってから、前記再生部は前記録画コンテンツの追っかけ再生を行う、
 テレビジョン装置。

【請求項2】

記録媒体に記録された蓄積コンテンツを再生する再生部を更に備え、

前記蓄積コンテンツの再生中に前記指示の入力を受けた場合、前記音声認識部が音声認識を行う間、前記再生部は前記蓄積コンテンツの再生を一時停止する、
請求項 1 に記載のテレビジョン装置。

【請求項 3】

前記制御部は、特定の音声認識結果を受けた場合、出力音量の設定値を前記指示の入力を受ける前の値に戻す請求項 1 に記載のテレビジョン装置。

【請求項 4】

ユーザの発話を検出する発話検出部を更に備え、
前記制御部は、前記指示の入力を受けてから所定時間を経過するまでの残り時間を画面に表示するとともに、

前記指示の入力を受けてから所定時間内に発話を検出されなかった場合、前記制御部は出力音量の設定値を前記指示の入力を受ける前の値に戻す、
請求項 1 に記載のテレビジョン装置。

【請求項 5】

前記音声入力部に入力された音から出力音声をキャンセルするエコーキャンセル部と、
前記テレビジョン装置のメインプロセッサの計算リソースを監視する計算リソース監視部と、
を備え、

前記制御部は、前記指示の入力を受けた時の計算リソースに応じて、出力音量の制御を行うか、前記エコーキャンセル部にエコーキャンセル処理を実行させるかを切り替える、
ことを特徴とする請求項 1 に記載のテレビジョン装置。

【請求項 6】

テレビジョン装置を操作するための遠隔操作装置であって、
音声を入力する音声入力部と、
前記音声入力部に入力される音量に基づいて、前記テレビジョン装置の出力音量レベルを推定する音量推定部と、

音声認識を開始する指示の入力をユーザから受ける指示入力部と、
前記指示の入力を受けた後のユーザの発話に対して音声認識を行って音声認識結果を得る音声認識部と、

前記音声認識結果に対応付けられた操作コマンドを表す信号を前記テレビジョン装置に送信する送信部と、
を有し、

前記送信部は、前記出力音量レベルの推定値が閾値以上である場合、前記音声認識部が音声認識を行う間の前記テレビジョン装置の出力音量の設定値を一時的に所定値に変更するための特定の操作コマンドを前記テレビジョン装置に送信し、

前記テレビジョン装置は 放送番組を録画して録画コンテンツを得る録画部と、
録画コンテンツを再生する再生部とを更に備え、
前記放送番組の視聴中に前記指示の入力を受けた場合、前記録画部は前記放送番組の録画を開始し、

前記音声認識部による音声認識処理が終わってから、前記再生部は前記録画コンテンツの追っかけ再生を行う遠隔操作装置。

【発明の詳細な説明】

【技術分野】

【0001】

本発明の実施形態は、音声で操作可能なテレビジョン装置及び遠隔操作装置に関する。

【背景技術】

【0002】

ユーザの発話を音声認識して、機器の操作に使う技術がある。操作対象の機器が音声を発する機器である場合、その機器が発する音声が音声認識の上では雑音となる。機器が発する音声と話者が発した音声が混在して入力された信号から、機器が発する音声をキャン

10

20

30

40

50

セルするエコーキャンセル技術を使って音声認識精度を向上させる技術がある。しかし、エコーキャンセルのための演算処理が必要となるため、処理能力に制約がある機器では実現が困難な場合がある。

【0003】

音声認識中に機器が発する音声をミュートする装置がある。音声認識中は機器の音声がないため、その影響を受けることなく音声認識できる。しかし、操作対象の機器がテレビ受像機である場合、視聴者は音声認識中に放送された音声を聞くことができない。

【先行技術文献】

【特許文献】

【0004】

【特許文献1】特開2009-109536

【特許文献2】特開平11-15494

【発明の概要】

【発明が解決しようとする課題】

【0005】

発明が解決しようとする課題は、少ない処理量で音声認識の精度を保ちつつ、音声での操作によって視聴が妨げられにくいテレビ受像機を提供することである。

【課題を解決するための手段】

【0006】

実施形態のテレビジョン装置は、音声を入力する音声入力部と、音声認識を開始する指示の入力をユーザから受ける指示入力部と、前記指示の入力を受けた後のユーザの発話に対して音声認識を行って音声認識結果を得る音声認識部と、前記音声認識結果に対応付けられた操作コマンドを実行する制御部と放送番組を録画して録画コンテンツを得る録画部と、録画コンテンツを再生する再生部とを有する。前記制御部は、前記指示の入力を受けた時の出力音量の設定値が閾値以上である場合、前記音声認識部が音声認識を行う間の出力音量の設定値を一時的に前記閾値以下の値に変更し、前記放送番組の視聴中に前記指示の入力を受けた場合、前記録画部は前記放送番組の録画を開始し、前記音声認識部による音声認識処理が終わってから、前記再生部は前記録画コンテンツの追っかけ再生を行う。

【図面の簡単な説明】

【0007】

【図1】第1の実施形態のテレビジョン装置のブロック図。

【図2】第1の実施形態のテレビジョン装置の外観の一例を示す模式図。

【図3】第1の実施形態のテレビジョン装置の外観の一例を示す模式図。

【図4】第1の実施形態のテレビジョン装置のハードウェア構成を示す図。

【図5】第1の実施形態の音声認識装置100の処理のフローチャート。

【図6】第1の実施形態にテレビ受像機110の処理のフローチャート。

【図7】第1の実施形態の音量変更の設定値の一例。

【図8】第1の実施形態の音量変更の設定値の一例。

【図9】第1の実施形態の音量変更の設定値の一例。

【図10】第1の実施形態の音量変更の設定値の一例。

【図11】第1の実施形態のタイムアウト時間の表示例の模式図。

【図12】第1の実施形態のLED点灯パターンの一例を示した図。

【図13】第1の実施形態の音声認識コマンドと操作コマンドの対応関係の一例。

【図14】第1の実施形態の音声認識コマンドと操作コマンドの対応関係の一例。

【図15】第2の実施形態のテレビ受像機110の処理を示すフローチャート。

【図16】第3の実施形態のテレビジョン装置のブロック図。

【図17】第3の実施形態のテレビジョン装置の外観の一例を示す模式図。

【図18】第3の実施形態の音声認識装置100の処理のフローチャート。

【図19】第3の実施形態のテレビ受像機110の処理のフローチャート。

【図20】第4の実施形態のテレビジョン装置のブロック図。

10

20

30

40

50

【図 2 1】第 4 の実施形態のテレビ受像機 1 1 0 の処理のフローチャート。

【発明を実施するための形態】

【0008】

以下、実施形態について図面を参照して詳細に説明する。本願明細書と各図において、既出の図に関して前述したものと同様の要素には同一の符号を付して詳細な説明は適宜省略する。

【0009】

(第 1 の実施形態)

図 1 は本実施形態のテレビジョン装置 1 0 の構成を示すブロック図である。ユーザは音声でテレビジョン装置 1 0 を操作することができる。テレビジョン装置 1 0 は音声認識装置 1 0 0 およびテレビ受像機 1 1 0 を有する。音声認識装置 1 0 0 はテレビ受像機 1 1 0 に内蔵されるか、テレビ受像機 1 1 0 と無線または有線で接続される。

【0010】

図 2 は、音声認識装置 1 0 0 がテレビ受像機 1 1 0 に内蔵された場合のテレビジョン装置 1 0 の外観の一例を示す模式図である。図 3 は、音声認識装置 1 0 0 がテレビ受像機とケーブルで接続された場合のテレビジョン装置 1 0 の外観の一例を示す模式図である。図 2 及び図 3 のいずれの場合もテレビジョン装置 1 0 の構成のブロック図は図 1 である。尚、図 2、図 3 ではマイク 2 0 4 がともに 2 個設置された場合の外観を示しているが、1 個であっても、3 個以上あっても構わなく、その設置場所も図示された限りではない。

【0011】

第 1 の実施形態に係る音声認識装置 1 0 0 は、マイク 1 0 1 と、マイクからの音声を取り込む音声入力部 1 0 2 と、予め定められたユーザからの音声認識開始の合図を検出する音声認識開始検出部 1 0 3 と、ユーザの発話の有無を検出する発話検出部 1 0 4 と、無音区間検出により音声認識終了を検出する音声認識終了検出部 1 0 5 と、エコーキャンセル部 1 0 6 と、音声入力部 1 0 2 から入力された音声を認識する音声認識部 1 0 7 と、音声認識結果に基づいて所定の信号をテレビ受像機 1 1 0 に送信する信号送信部 1 0 8 を有する。音声認識開始検出部 1 0 3 は、ユーザからの合図により音声認識の入力が開始されてからのユーザ発話の有無を検出する。エコーキャンセル部 1 0 6 は、マイク 1 0 1 を経由して音声入力部 1 0 2 に入力されるテレビ受像機 1 1 0 のスピーカ 1 1 5 から発せられる音声をキャンセルする。

【0012】

テレビ受像機 1 1 0 は、信号送信部 1 0 8 から送られてきた信号に基づいてテレビ音量の制御をはじめ所望のテレビ操作を実行するテレビ制御部 1 1 1 と、テレビ受像機 1 1 0 のメインプロセッサの計算リソースをモニタする計算リソース監視部 1 1 2 と、映像再生部 1 1 3 と、放送されている番組を録画する録画部 1 1 4 と、視聴している番組の音を出力するスピーカ 1 1 5 と、視聴している番組の映像を表示する表示部 1 1 6 を有する。映像再生部 1 1 3 は、放送中の番組コンテンツ、録画された番組コンテンツ、又は、記録媒体に記録された映像コンテンツを再生する。記録媒体として、例えば DVD (Digital Versatile Disc) や BD (Blu-ray Disc) が挙げられる。

【0013】

図 4 は音声認識装置 1 0 0 およびテレビ受像機 1 1 0 のハードウェア構成の例を示す。図 4 のハードウェアは、装置全体を制御する CPU (Central Processing Unit) 等の制御部 1 3 0 と、各種データや各種プログラムを記憶する ROM (Read Only Memory) や RAM (Random Access Memory) 等の記憶部 1 3 1 と、各種データや各種プログラムを記憶する HDD (Hard Disk Drive) や CD (Compact Disk) ドライブ装置等の外部記憶部 1 3 2 と、ユーザの指示入力を受け付ける操作部 1 3 3 と、外部装置との通信を制御する通信部 1 3 4 と、これらを接続するバス 1 3 5 とを備える。さらに、音声認識装置 1 0 0 には、音声が入力されるマイクなどの音声入力部 1 3 6 が有線又は無線により接続される。なお、図 4 のハードウェアの構成は一例である。音声認識装置 1 0 0 及びテレビ受像機 1 1 0 の一部ないし全部の構成が LSI や ASIC などの専用の半導体集積回路で実現され

10

20

30

40

50

ても構わない。

【 0 0 1 4 】

図 5 は、第 1 の実施形態に係る音声認識装置 1 0 0 の処理動作を示すフローチャートである。図 6 は、第 1 の実施形態に係るテレビ受像機 1 1 0 の処理動作を示すフローチャートである。次に、第 1 の実施形態に係る音声認識装置 1 0 0 およびテレビ受像機 1 1 0 の処理動作について、図 5、図 6 のフローチャートを参照して説明する。

【 0 0 1 5 】

まず、音声認識開始検出部 1 0 3 は、ユーザからの音声認識開始指示の入力を待つ（ステップ S 1）。ユーザからの音声認識開始指示として、予め定められた音を用いる。例えば、手を連続して 2 回叩いた音を指示として用いる。この場合、マイク 1 0 1 から入力された音声の中から、手を連続して 2 回叩いた音を検出する。

10

【 0 0 1 6 】

別の例として、ユーザが発した特定の言葉を用いることもできる。合図に用いる言葉を認識するための合図認識用辞書と、テレビ操作コマンドの言葉を認識するためのコマンド認識用辞書とを用意しておく。普段、音声認識部 1 0 7 は合図認識用辞書を用いて音声認識を行う。合図となる言葉が認識された時に、音声認識部 1 0 7 は音声認識用の辞書をコマンド認識用辞書に切り替える。さらに別の例として、リモコン 2 0 5 に音声認識開始ボタンを設け、そのボタン押下に対応したリモコン信号をリモコン受信部 2 0 3 から受信したことを音声認識開始の合図としても良い。これら以外の合図としても良い。またこれらの合図を一つだけ受理するようにしても良いし、複数の合図により音声認識開始を指示できるようにしても良い。

20

【 0 0 1 7 】

音声認識開始検出部 1 0 3 が音声認識開始の合図を検出した場合に（ステップ S 2）、信号送信部 1 0 8 は、テレビ受像機 1 1 0 のテレビ制御部 1 1 1 に音声認識開始信号を送信する（ステップ S 3）。このとき、音声認識が開始されたことをユーザにフィードバックするために、LED（発光ダイオード）の点灯や OSD（On-Screen Display）でその旨表示しても良い。

【 0 0 1 8 】

テレビ受像機 1 1 0 は、まず音声認識装置 1 0 0 の信号送信部 1 0 8 からの信号の待ち受け状態となる（ステップ S 1 0 1）。信号送信部 1 0 8 から何らかの信号を受信すると、テレビ制御部 1 1 1 は音声認識開始コマンドかどうかを判断する（ステップ S 1 0 2）。受信したコマンドが音声認識コマンドである場合、テレビ受像機 1 1 0 の映像再生部 1 1 3 は、現在表示している映像が、放送コンテンツであるか、蓄積コンテンツあるかを判断する（ステップ S 1 0 3）。放送コンテンツとは、地上波デジタル放送、BS デジタル放送、CS デジタル放送及び CATV など放送されている映像である。蓄積コンテンツとは、録画部 1 1 4 に録画された番組、及び、DVD や BD 等のメディアに記録された映像である。

30

【 0 0 1 9 】

視聴している映像が放送波である場合、計算リソース監視部 1 1 2 は、テレビ受像機 1 1 0 の制御部 1 3 0 内の CPU の計算負荷を計測し（ステップ S 1 0 4）、計算負荷が所定の閾値より大きい小さいかを判定する（ステップ S 1 0 5）。この際、計算負荷が CPU の全リソースの内どのくらいの割合を占有しているかを基準に判定しても良いし、現在テレビ受像機 1 1 0 で実行する処理ごとに計算量を定義しておき、現在実行している処理の計算量の総和で判定しても良い。エコーキャンセルに要する計算量を事前に調べておき、CPU でエコーキャンセル処理を実行する余裕があるかどうかを基準に閾値は設定される。したがって、CPU の計算負荷が閾値より小さければ、CPU にエコーキャンセル処理を行う余裕があることを意味する。計算負荷が所定の閾値より小さい場合はエコーキャンセル処理を実行し（ステップ S 1 0 6）、音声認識部 1 0 7 は音声認識の対象となる音声信号の入力（音声認識入力）を開始する（ステップ S 4）。この場合は、テレビ制御部 1 1 1 は音量変更を行わない。

40

50

【 0 0 2 0 】

計算負荷が所定の閾値以上である場合、テレビ制御部 1 1 1 は、現在のテレビ音量の設定値を読み取る（ステップ S 1 0 7）。その設定値が所定の閾値より大きい小さいかでテレビ音量の変更動作を変更する。図 7、図 8、図 9 及び図 1 0 に音量レベルの変更動作設定の例を示す。なお、これらの図では、音量レベルの数字が大きくなるほど音量が大きくなるものとする。

【 0 0 2 1 】

図 7 の設定例では、テレビ制御部 1 1 1 は音量レベルが第 1 の閾値以上であれば音声をミュートし、音量レベルが第 1 の閾値未満であれば音量レベルを変更しない。なお「音量レベルを変更しない」とは「音量レベルを保つ」と言い換えてもよい。

10

【 0 0 2 2 】

図 8 の設定例では、テレビ制御部 1 1 1 は音量レベルが第 1 の閾値以上であれば音量レベルをある一定のレベルに設定し、音量レベルが第 1 の閾値未満であれば音量レベルを変更しない。例えば、音量レベルが 2 0 以上の場合は音量レベルを一定値（例えば 2 0 以下の値）に設定し、音量レベルが 2 0 未満の場合には音量レベルを変更しない。

【 0 0 2 3 】

図 9 の設定例では、第 1 の閾値と第 2 の閾値（第 1 の閾値 > 第 2 の閾値）が用いられる。音量レベル第 1 の閾値以上の場合、テレビ制御部 1 1 1 は音声をミュートする。音量レベル第 2 の閾値以上第 1 の閾値未満の場合、テレビ制御部 1 1 1 は音量レベルを一定の値に下げる。音量レベル第 2 の閾値未満の場合、テレビ制御部 1 1 1 は音量レベルを変更しない。

20

【 0 0 2 4 】

図 1 0 の設定例では、第 1 の閾値と第 2 の閾値（第 1 の閾値 > 第 2 の閾値）が用いられる。第 1 の閾値以上の場合、テレビ制御部 1 1 1 は音声をミュートする。第 2 の閾値以上第 1 の閾値未満の場合、テレビ制御部 1 1 1 は音量レベルを現在値の半分のレベルに下げる。第 2 の閾値未満の場合、テレビ制御部 1 1 1 は音量を変更しない。例えば、音量レベルが 4 0 以上の場合はミュート、音量レベルが 2 0 以上 4 0 未満の場合は 1 / 2 ミュート、音量レベルが 2 0 未満の場合は音量を変更しないという設定である。もちろん、他の設定でも構わない。

【 0 0 2 5 】

図 7、図 8、図 9 及び図 1 0 では、第 1 の閾値として「 2 0 」が用いられ、第 2 の閾値として「 4 0 」を用いられたが、この数値に限定されるものではない。上述の第 1、第 2 の閾値を他の音量レベルに変更しても構わない。一般的には事前に音声認識性能評価を行い、その結果に応じて音声認識性能を確保できる程度に適宜設定される。上述の例の第 1 の閾値「 2 0 」や、第 2 の閾値「 4 0 」という数値そのものは、音声認識装置 1 0 0 及びテレビ受像機 1 1 0 の構成・性能に応じて適宜変更される。

30

【 0 0 2 6 】

このようにしてテレビ音量に応じた音量変更を実行した後、音声認識部 1 0 7 は音声認識入力を実行する（ステップ S 4）。

【 0 0 2 7 】

一方、蓄積されたコンテンツを視聴している場合、映像再生部 1 1 3 は再生中の映像を一時停止させ（ステップ S 1 0 9）、音声認識部 1 0 7 は音声認識入力を実行する（ステップ S 4）。蓄積されたコンテンツとは、例えば録画部 1 1 4 に録画された番組や、DVD、BD 等の記録メディアに記録された映像である。

40

【 0 0 2 8 】

音声認識装置 1 0 0 の発話検出部 1 0 4 はユーザの発話が始まったかどうかを検出する。ユーザが誤って音声認識開始の合図をした場合や、音声認識開始検出部 1 0 3 が音声認識開始の合図を誤検出した場合に、自動的に元の状態へ復帰させるためのタイムアウトを設定しておくが良い。さらに、図 1 1 の表示 1 1 0 1 のようにタイムアウトまでの時間を OSD 表示しても良いし、図 1 2 に示すパターンのように LED の点灯や点滅で表示し

50

ても良い。また、発話検出部 104 がユーザの発話を検出する前後で、LED点滅、OSD表示を変更すると、ユーザは自分の発話が受理されているかがわかりやすい。図12はLEDによるステータス表示の一例である。図12のグラフの横軸は時間であり、縦軸はLEDの点灯・消灯の状態を表す。この例では、時刻 t_1 で音声認識開始合図が検出されるとLEDが点灯し、時刻 t_2 で発話を検出されるとLEDが点滅し、時刻 t_3 で音声認識が終了するとLEDが消灯する。

【0029】

音声認識終了検出部 105 は、音声認識を終了するか否かを判定する（ステップ S5）。例えば「無音区間が一定時間以上継続していること」は本実施形態の音声認識終了条件の一つである。音声認識部 107 は音声認識を実行し音声認識結果を得る（ステップ S6）。信号送信部 108 は音声認識結果に応じてテレビ受像機 110 の操作コマンドをテレビ音量制御部 111 に送信する（ステップ S7）。

10

【0030】

ここでは、チャンネル切替、音量変更、入力切り替え、画面モード切替など、あらかじめ特定の音声コマンド（音声認識結果）と対応づけられた操作コマンドを送信する。操作コマンドと音声コマンドとの対応付けの例を図13のテーブル1300と図14のテーブル1400に示す。この際、ユーザが誤って音声認識開始の合図をしてしまった場合や、音声認識開始検出部 103 が音声認識開始の合図を語検出した場合に元の状態に復帰するための音声コマンド（図13のコマンド1301）を用意しておくが良い。

【0031】

テレビ受像機 110 は、音声認識開始コマンド以外の操作コマンドを受信した場合（ステップ S102 - No）、キャンセルコマンドかどうかを判断する（ステップ S110）。キャンセルコマンドである場合（ステップ S110 - Yes）、テレビ操作を実行せずに音声認識開始前の状態に戻す（ステップ S112）。キャンセルコマンドでない場合（ステップ S111 - No）、受信した操作コマンドに対応づけられたテレビ操作を実行し（ステップ S111）、音声認識開始前の音量設定に戻す（ステップ S112）。

20

【0032】

上述したように、本実施形態のテレビは、音声認識開始前のテレビ音量に応じて、音声認識処理の間のテレビ音量を一時的に制御する。これにより、少ない演算量で精度良い音声認識をしつつ、音声での操作によって視聴が妨げられにくいテレビが実現される。

30

【0033】

また、本実施形態のテレビは、蓄積コンテンツを再生している場合、音声認識中に再生を一時停止させる。これにより、音声での操作中に、蓄積コンテンツを不完全な状態で視聴することが避けられる。

【0034】

（第2の実施形態）

図面を用いて第2の実施形態のテレビを説明する。第1の実施形態と同じ処理・構成については同じ符号を付して説明を省略し、第1の実施形態と異なる部分を説明する。図15は、本実施形態のテレビ受像機 110 が行う処理のフローチャートである。

【0035】

テレビ受像機 110 は、音声認識開始コマンドを受信した後、現在の視聴メディアに応じて処理動作を変える（ステップ S103）。現在の視聴メディアが放送である場合、テレビ制御部 111 は画面を静止させるとともに音声をミュート状態にする（ステップ S201）。その後、すぐに録画部 114 はその放送の録画を開始する（ステップ S202）。

40

【0036】

テレビ制御部 111 は、音声認識終了後に音声認識結果に基づく操作コマンドを受信し、操作コマンドに対応するテレビ操作を実行する（ステップ S111）。制御部 111 は（1）音声認識開始前の視聴メディアが放送であるか、（2）ステップ S201 の処理によって画面が静止し、かつ、音声がミュート状態になっているか、（3）音声認識開始前

50

の視聴メディアの録画が開始されているか、並びに、(4) テレビ制御部 111 によって実行されたテレビ操作が放送波のチャンネル変更でないか、の4つの条件を満たすかを判定する(ステップ S203)。(1)~(4)のすべての条件を満たす場合は、静止した画面からの追っかけ再生を開始する(ステップ S204)。典型的には、チャンネル変更でない操作(例えば音量変更)が行われた場合である。

【0037】

一方、(1)~(4)のいずれか一つでも条件を満たさない場合には、追っかけ再生を行わずに、制御部 111 は音声認識開始前の音量設定に復帰させる(ステップ S112)。ステップ S202 で録画処理が実行され、その後、視聴チャンネルの変更が行われた場合は録画を停止しても構わない。録画を停止した場合、録画したデータは消去しても構

10

【0038】

本実施形態のテレビは、音声をミュートした状態で音声認識できるので、少ない演算量で精度良く音声認識できる。また、音声認識中の放送内容が録画され、音声認識後に追っかけ再生されるので、音声でテレビを操作しても視聴が妨げられにくくなる。

【0039】

(第3の実施形態)

図面を用いて第3実施形態のテレビを説明する。第1及び第2の実施形態と同じ処理及び構成については同じ符号を付して説明を省略する。第1及び第2の実施形態と異なる処理及び構成を説明する。

20

【0040】

図16は、本実施形態の音声認識装置 100 およびテレビ受像機 110 の構成を示すブロック図である。図17は、第3の実施形態に係る音声認識装置 100 およびテレビ受像機の外観の一例を示す模式図である。

【0041】

図17に示すように、本実施形態の音声認識装置 100 (210) とテレビ受像機 110 (200) は離れた位置にある。本実施形態の音声認識装置 100 はテレビ受像機 110 を離れた場所から操作するためのリモコン装置(リモートコマンダー)である。本実施形態の音声認識装置 100 はテレビ受像機 110 に無線(例えば、赤外線パルスや電波)で操作コマンドを送信する。

30

【0042】

本実施形態の音声認識装置 100 は、音声認識装置 100 が置かれた位置での環境音を推定するために、テレビ音量推定部 120 を備える。テレビ音量計測部 120 は、音声入力部 102 に入力された環境音の過去一定時間の音量の平均値からテレビ音量を推定する。

【0043】

本実施形態の信号送信部 108 は、テレビ音量推定部 120 で推定されたテレビ音量に応じて音声認識中のテレビ受像機 110 の音量レベルを変更する。信号送信部 108 は推定された音量レベルに基づいて音声認識中の音量レベルを求める。推定された音量レベルと音声認識中の音量レベルとの対応関係として、例えば図7、図8、図9及び図10に示した設定例を用いることができる。もちろん、他の設定例を用いても構わない。信号送信部 108 は求められた音量レベルに設定するための操作コマンドをテレビ受像機 110 へ送信する。信号送信部 108 は音量レベルを下げる操作コマンドを繰り返し送信しても構わないし、音量レベルの値を直接指定する操作コマンド(ダイレクトコード)を送信しても構わない。また、音量レベルを半分にする(1/2ミュート)のような特別な操作コマンドを送信しても構わない。音声認識を行う間の音量レベルを、ある一定の水準よりも低くすることができれば他の操作コマンドでも構わない。

40

【0044】

図18は、第3の実施形態に係る音声認識装置 100 の処理動作を示すフローチャートである。図19は、第3の実施形態に係るテレビ受像機 110 の処理動作を示すフローチ

50

ャートである。

【0045】

音声認識開始検出部103が音声認識開始を検出すると、テレビ音量推定部120は音声入力部102に入力された環境音の過去一定時間の音量の平均値からテレビ音量を推定する(ステップS10)。推定されたテレビ音量に応じて、信号送信部108は音声認識中のテレビ音量を変更するための操作コマンドを送信する(ステップS11)。その後、音声認識部107は音声認識を行って音声認識結果を得る(ステップS4、ステップS5、ステップS6)。信号送信部108は音声認識結果に基づく操作コマンド送信を実行し(ステップS7)、その後、ミュート解除コマンド等、音量を音声認識入力前の状態に戻すための操作コマンドを送信する(ステップS12)。

10

【0046】

上述したように、本実施形態のテレビは、音声認識開始前のテレビ音量に応じて、音声認識処理の間のテレビ音量を一時的に制御する。これにより、少ない演算量で精度良い音声認識をしつつ、音声での操作によって視聴が妨げられにくいテレビが実現される。また、本実施形態のテレビは音声認識部で計測されるテレビ音量に基づいて音声認識中のテレビ音量を制御する。これにより、音声認識のために必要な範囲でテレビ音量を制御できる。

【0047】

(第4の実施形態)

図面を用いて第4実施形態のテレビを説明する。第1～第3の実施形態と同じ処理及び構成については同じ符号を付して説明を省略する。第1～第3の実施形態と異なる処理及び構成を説明する。

20

【0048】

図20は本実施形態のテレビジョン装置10の構成を示すブロック図である。本実施形態のテレビジョン装置10は、音声認識装置100とテレビ受像機110を有する。音声認識装置100は、マイク101、音声入力部102、音声認識開始検出部103、発話検出部104、音声認識終了検出部105、音声認識部107、及び信号送信部108を有する。テレビ受像機110はテレビ制御部111、映像再生部113、スピーカー115及び表示部116を有する。

【0049】

図21は本実施形態に係るテレビ受像機110の処理動作を示すフローチャートである。本実施形態のテレビ受像機110は、蓄積コンテンツを視聴している場合にも放送コンテンツを視聴している場合にも、音声認識開始コマンドを受信した後に音量の制御を行う(ステップS107、ステップS108)。

30

【0050】

本実施形態のテレビは、音声認識開始前のテレビ音量に応じて、音声認識処理の間のテレビ音量を一時的に制御する。これにより、少ない演算量で精度良い音声認識をしつつ、音声での操作によって視聴が妨げられにくいテレビジョン装置が実現される。

【0051】

(変形例)

本発明は上述した各実施形態に限定されるものではない。各実施形態の各要素の具体的な構成に関しては、当業者が公知の範囲から適宜選択することにより本発明を同様に実施し、同様の効果を得ることができる限り、本発明の範囲に包含される。

40

【0052】

また、各実施形態のいずれか2つ以上の要素を技術的に可能な範囲で組み合わせたものも、本発明の要旨を包含する限り本発明の範囲に含まれる。一つの実施形態又は複数の実施形態を組み合わせたものを当業者が適宜設計変更したものも、本発明の要旨を包含する限り、本発明の範囲に属する。

【0053】

その他、本発明の思想の範疇において、当業者であれば、各種の変更例及び修正例に想

50

到し得るものであり、それら変更例及び修正例についても本発明の範囲に属するものと了解される。

【0054】

本発明の幾つかの実施形態を説明したが、これらの実施形態は、例として提示したものであり、発明の範囲を限定することは意図していない。これら新規な実施形態は、その他の様々な形態で実施されることが可能であり、発明の要旨を逸脱しない範囲で、種々の省略、置き換え、変更を行うことができる。これら実施形態やその変形は、発明の範囲や要旨に含まれるとともに、特許請求の範囲に記載された発明とその均等の範囲に含まれる。

【符号の説明】

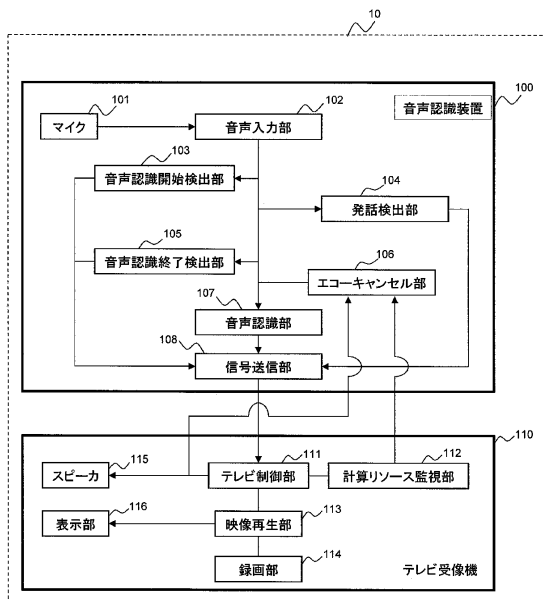
【0055】

10	テレビ	100	音声認識装置	101	マイク
102	音声入力部	103	音声認識開始検出部	104	発話検出部
105	音声認識終了検出部	106	エコーキャンセル部	107	音声認識部
108	信号送信部	110	テレビ受像機	111	テレビ音量制御部
112	計算リソース監視部	113	映像再生部	114	録画部
115	スピーカ	116	表示部	130	制御部
131	記憶部	132	外部記憶部	133	操作部
134	通信部	135	バス	136	音声入力部
200	テレビ受像機	201	フレーム	202	表示部
203	リモコン受信部	204	マイク	205	リモコン
210	音声認識装置				

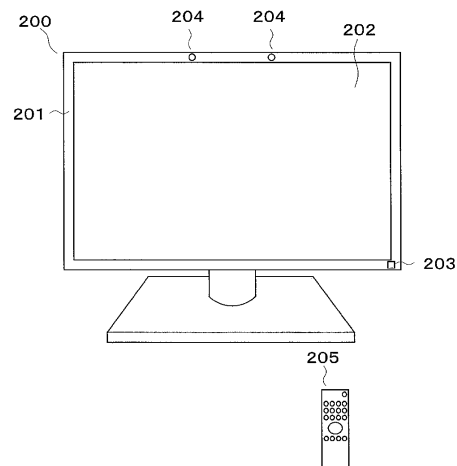
10

20

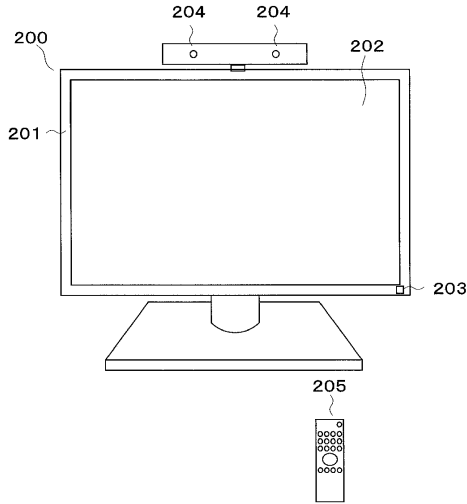
【図1】



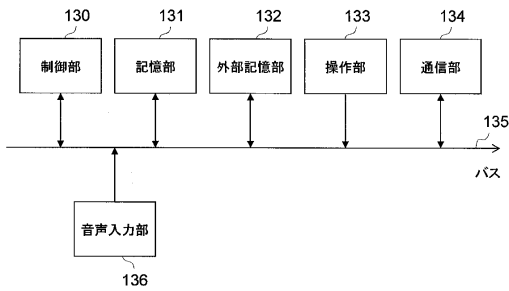
【図2】



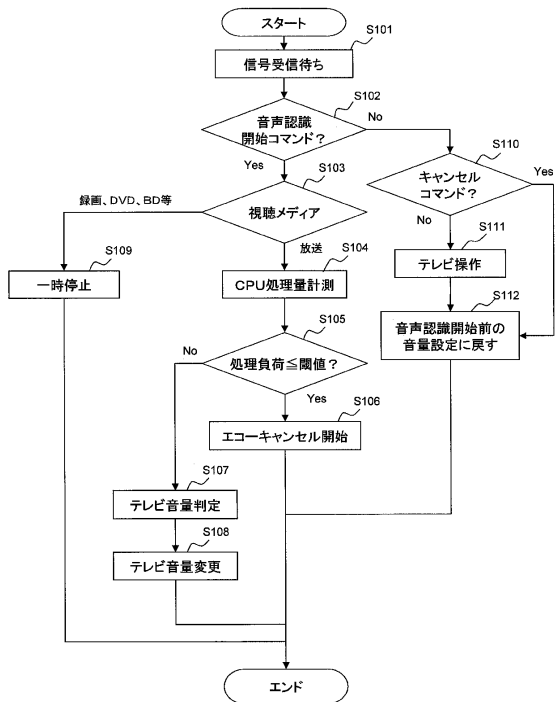
【図3】



【図4】



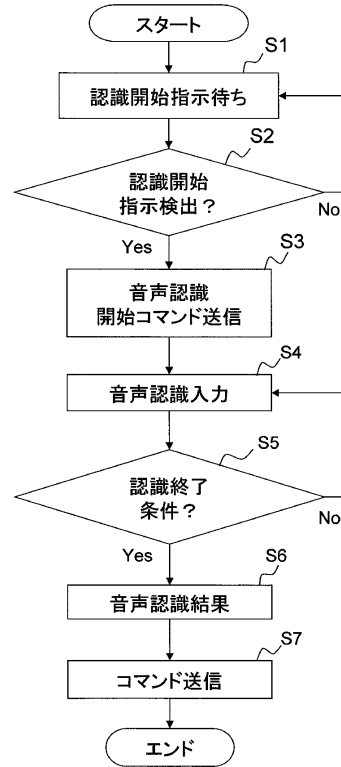
【図6】



【図7】

音声認識開始検出時の音量レベル	音声認識中の音量レベル
20以上	ミュート
20未満	音量変更無し

【図5】



【図8】

音声認識開始検出時の音量レベル	音声認識中の音量レベル
20以上	音量20
20未満	音量変更無し

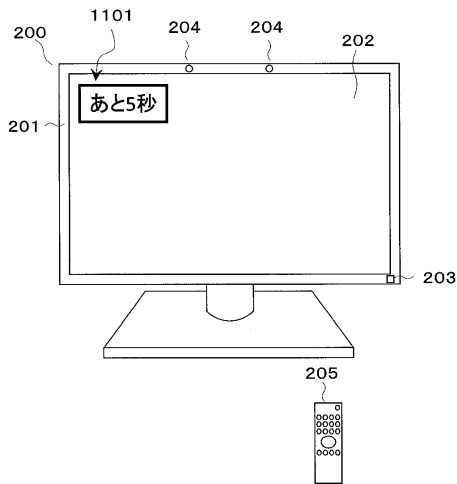
【図9】

音声認識開始検出時の音量レベル	音声認識中の音量レベル
40以上	ミュート
20以上40未満	音量20
20未満	音量変更無し

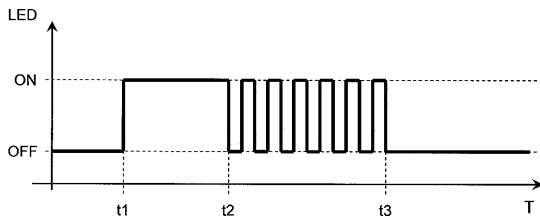
【図10】

音声認識開始検出時の音量レベル	音声認識中の音量レベル
40以上	ミュート
20以上40未満	1/2ミュート
20未満	音量変更無し

【図11】



【図12】



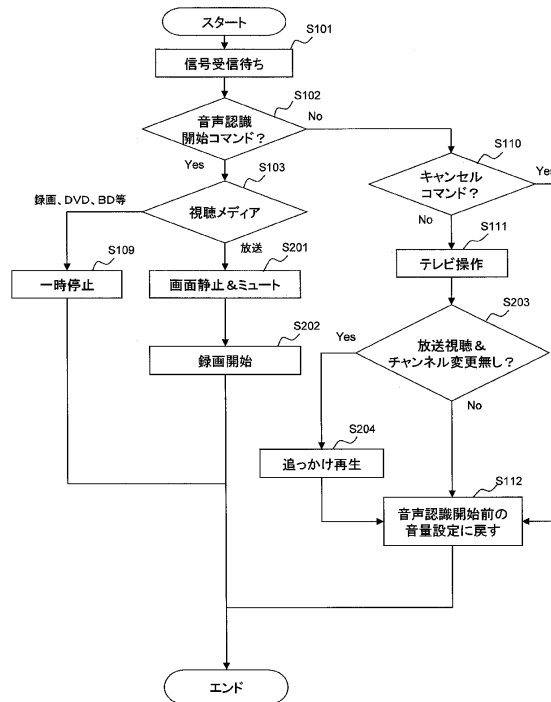
【図14】

テレビ操作の内容	認識言葉(日本語)	認識言葉(英語)
今日の天気予報を表示	てんきよほう	weather report
今日の天気予報を表示	きょうのてんき	today's weather
明日の天気予報を表示	あしたのてんき	tomorrow's weather
今週の天気予報を表示	こんしゅうのてんき	this week's weather
...		
スポーツニュースを表示	すぽーつ	sports
野球のニュースを表示	やきゅう	baseball
巨人軍の試合結果を表示	きょじん	giants
サッカーのニュースを表示	さっかー	soccer
...		
現在の為替相場を表示	かわせそうば	exchange rate
...		
最寄りのバス停の時刻表を表示	ばす	bus
最寄りの駅の時刻表を表示	でんしゃ	train
...		
...		

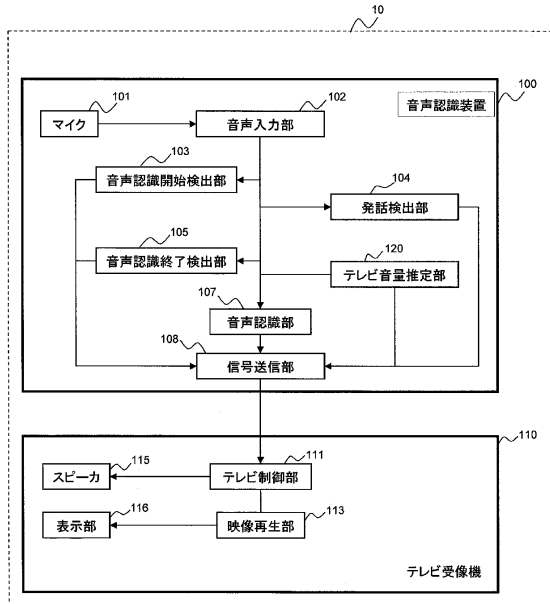
【図13】

テレビ操作の内容	認識言葉(日本語)	認識言葉(英語)
チャンネルを1にする	いっちゃん	channel one
チャンネルを1にする	えぬえちけー	nhk
チャンネルを2にする	にちゃん	channel two
チャンネルを2にする	きょういくてれび	etv
...		
音量を少し上げる	あつぷ	up
音量を大幅に上げる	あつぷあつぷ	up up
音量を少し下げる	だうん	down
音量を大幅に下げる	だうんだうん	down down
音量を小(規定値)にする	ちいさめ	low
音量を中(規定値)にする	ふつう	medium
音量を大(規定値)にする	おおきめ	loud
ミュートする	みゅーと	mute
...		
電源オフ	でんげんおふ	turn off
電源オフ	おふ	turn off
音声認識モードキャンセル	きゃんせーる	cancel

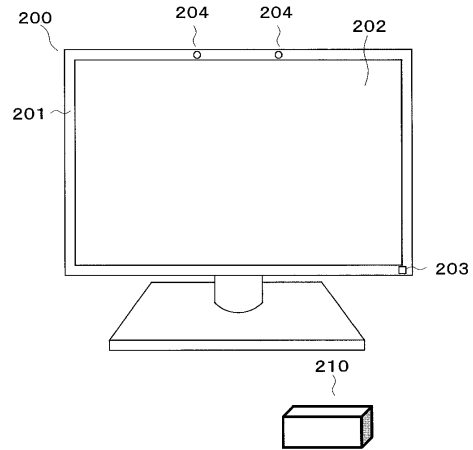
【図15】



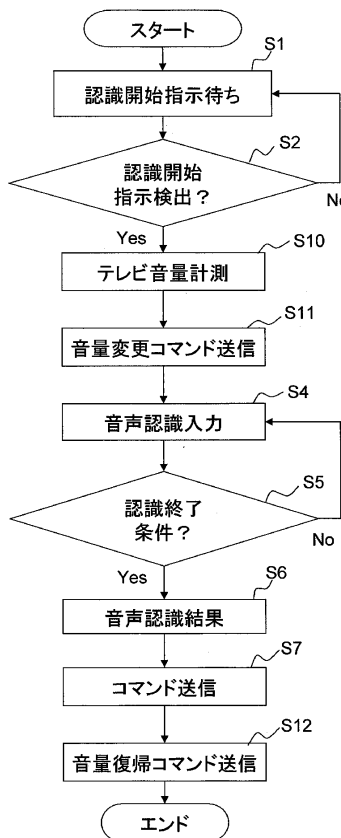
【図16】



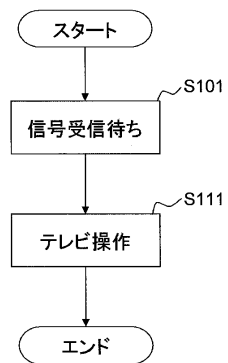
【図17】



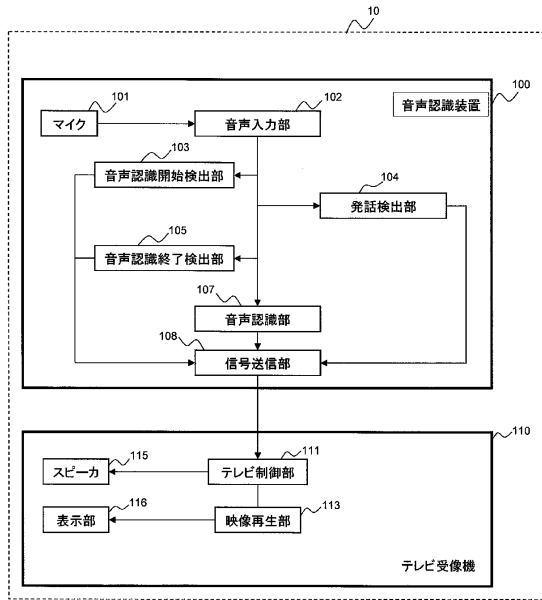
【図18】



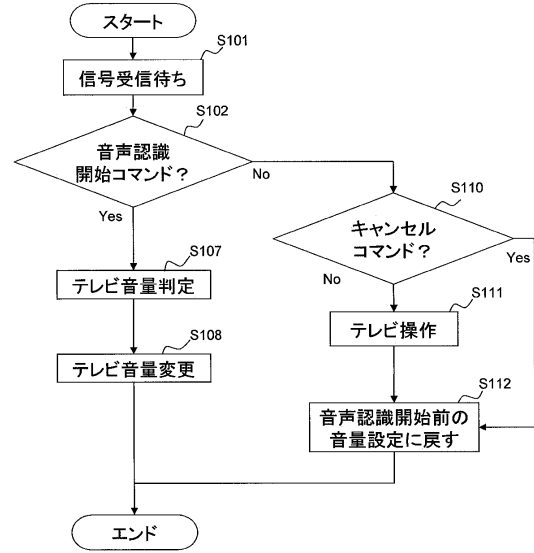
【図19】



【図20】



【図21】



フロントページの続き

- (72)発明者 鈴木 薫
東京都港区芝浦一丁目1番1号 株式会社東芝内
- (72)発明者 木田 祐介
東京都港区芝浦一丁目1番1号 株式会社東芝内

審査官 山下 剛史

- (56)参考文献 特開平3 - 203796 (JP, A)
特開2001 - 236090 (JP, A)
特開2006 - 65981 (JP, A)
特開2006 - 119520 (JP, A)
国際公開第2009 / 150776 (WO, A1)
米国特許出願公開第2007 / 0014420 (US, A1)

- (58)調査した分野(Int.Cl., DB名)
- | | |
|------|-------------------|
| G10L | 15 / 00 - 15 / 28 |
| H04N | 7 / 14 - 7 / 173 |