

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4675840号
(P4675840)

(45) 発行日 平成23年4月27日(2011.4.27)

(24) 登録日 平成23年2月4日(2011.2.4)

(51) Int.Cl.		F I		
G 1 0 L	15/04	(2006.01)	G 1 0 L	15/04 3 0 0 B
G 1 0 L	15/28	(2006.01)	G 1 0 L	15/28 2 3 0 K
G 1 0 L	15/00	(2006.01)	G 1 0 L	15/00 2 0 0 F

請求項の数 13 (全 14 頁)

(21) 出願番号	特願2006-179261 (P2006-179261)	(73) 特許権者	000006013 三菱電機株式会社 東京都千代田区丸の内二丁目7番3号
(22) 出願日	平成18年6月29日(2006.6.29)	(74) 代理人	100085198 弁理士 小林 久夫
(65) 公開番号	特開2008-9120 (P2008-9120A)	(74) 代理人	100098604 弁理士 安島 清
(43) 公開日	平成20年1月17日(2008.1.17)	(74) 代理人	100061273 弁理士 佐々木 宗治
審査請求日	平成20年5月30日(2008.5.30)	(74) 代理人	100070563 弁理士 大村 昇
		(74) 代理人	100087620 弁理士 高梨 範夫

最終頁に続く

(54) 【発明の名称】 リモートコントローラ並びに家電機器

(57) 【特許請求の範囲】

【請求項1】

周期的な所定期間を示す区間を複数有する複数区間毎にこの複数区間内の音声を取り込み、FFTにより周波数成分に変換し、所定の周波数帯における各周波数成分のレベルの積算量を一つ前の区間のそれと比較して差分を算出し、この差分が音声認識開始用の閾値を超えた第1の開始区間及び次に前記差分が音声認識終了用の閾値を超えた第1の終了区間を求め、さらに前記各区間毎にこの区間内の音声を取り込み、FFTにより周波数成分に変換し、所定の周波数帯における各周波数成分のレベルの積算量を一つ前の区間のそれと比較して差分を算出し、この差分が前記音声認識開始用の閾値を超えた第2の開始区間から次に前記差分が前記音声認識終了用の閾値を超えた第2の終了区間を求め、前記第1の開始区間と前記第2の開始区間に基づき開始区間を推定し、前記第1の終了区間及び前記第2の終了区間に基づき終了区間を推定し、推定された開始区間から推定された終了区間までの一連の複数区間の音声信号を出力する音響分析部と、

この音響分析部が出力した一連の複数区間の音声信号レベルを基に音声認識アルゴリズムを適用して前記複数区間の各区間毎に音声特徴ベクトルの抽出や他の補足計算を行う特徴量・補足計算部と、

この特徴量・補足計算部によって抽出された特徴ベクトルと予め保有するデータベースの語句とに基づいて音声認識を行う音声認識部と、

この音声認識部が認識した結果を操作指令として外部機器に送信する通信部とを備えたことを特徴とするリモートコントローラ。

【請求項 2】

前記音響分析部は、前記差分が前記音声認識開始用の閾値を超えた区間を音声開始区間とし、所定の基準により少なくとも前記音声認識開始用の閾値を超えた区間の前後の区間を単語発声開始時点とし、前記差分が前記音声認識終了用の閾値を超えた区間を音声終了区間とし、所定の基準により少なくとも前記音声認識終了用の閾値を超えた区間の前後の区間を単語発声終了時点とすることを特徴とした請求項 1 記載のリモートコントローラ。

【請求項 3】

前記音響分析部は、区間毎でなく複数区間（以下、周期という）毎にこの周期内の音声を取り込み、周波数成分に変換し、所定の周波数帯における各周波数成分のレベルの積算量を一つ前の周期のそれと比較して差分を算出し、この差分が前記音声認識開始用の閾値を超えた区間から一連の複数区間の音声信号を出力することを特徴とする請求項 1 記載のリモートコントローラ。

10

【請求項 4】

前記音響分析部は、音声入力区間を検出後、前記音声認識開始の閾値及び前記音声認識終了の閾値を超える差分が存在した周波数帯域だけについて前記音声入力区間内の各区間毎の周波数成分を FFT 逆変換した結果を音声信号として出力することを特徴とする請求項 1 記載のリモートコントローラ。

【請求項 5】

前記音響分析部は、音声入力区間を検出後、変化があった周波数帯域の周波数成分の重み付けを重くしてこの周波数成分を再構成し、前記音声入力区間内の各区間毎の周波数成分を FFT 逆変換した結果を音声信号として出力することを特徴とする請求項 1 記載のリモートコントローラ。

20

【請求項 6】

前記音響分析部は、音声帯域の周波数成分で予め定めた上限値よりも大きいレベルの周波数成分を除去することを特徴とした請求項 1 記載のリモートコントローラ。

【請求項 7】

前記音響分析部は、音声帯域の周波数成分で予め定めた上限値よりも大きいレベルの周波数成分の重みを軽くしたことを特徴とした請求項 1 記載のリモートコントローラ。

【請求項 8】

前記音響分析部は、音声帯域の周波数成分で予め定めた上限値よりも大きいレベルの周波数成分を、隣接する周波数の成分のレベルの平均値に置き換えたことを特徴とした請求項 1 記載のリモートコントローラ。

30

【請求項 9】

前記音声認識部は、音声入力開始区間検出後、音声認識部のアルゴリズムに合わせて、単語発生区間の終了を待たずに並列処理にて順次処理することを特徴とする請求項 1 ~ 8 のいずれかに記載のリモートコントローラ。

【請求項 10】

前記音響分析部は、事前に登録した音声や定期的（周期的）に取り込んだ音声区間の解析結果を基に周波数帯、差分の閾値を決定し、この閾値を用いて入力した音声が発音雑音なのかスパイクノイズなのかを判断することを特徴とする請求項 1 記載のリモートコントローラ。

40

【請求項 11】

プロセッサと、音声取込部と、音声出力部と、ボタンと、表示部と通信部とを備え、
前記音声取込部は、前記ボタンから信号を受信すると外部から最初の音声を取り込み、
前記プロセッサは前記音響分析部、特徴量・補足計算部及び音声認識部を実現し、前記音声取込部が取り込んだ前記最初の音声を認識し、認識した結果を外部のユーザーに伝えるべく音声出力部へ出力し、前記ボタンあるいは前記音声取込部から許可の旨の信号あるいは許可の旨の音声を受信すると、音声認識し、前記最初の音声を認識した結果を操作指令として前記通信部を介して外部の機器に送信することを特徴とする請求項 1 ~ 10 のいずれかに記載のリモートコントローラ。

50

【請求項 1 2】

データベースを備え、

前記プロセッサは前記最初の音声を認識した結果を学習し前記データベースに反映することを特徴とする請求項 1 1 記載のリモートコントローラ。

【請求項 1 3】

通信部を備え、

請求項 1 ~ 1 2 のいずれかに記載のリモートコントローラから前記通信部を介して操作指令を受信してこの操作指令に基づいて動作することを特徴とする家電機器。

【発明の詳細な説明】

【技術分野】

10

【0001】

この発明は、空調機などの家電機器を操作するリモートコントローラ並びに家電機器に関するものである。

【背景技術】

【0002】

家電機器はユニバーサルデザイン化を目指し、音声認識を用いるものが存在する。空調機などの場合、機器自体の距離が遠いため、リモコンで音声認識を行う技術が紹介開示されている。

【0003】

例えば、ユーザーがボタン 1 個の動作と 5 種類の日常単語を発声するだけで、その後引き続き所定時間のみ音声認識を動作させ、対話方式で機器を動作させ、複数個の煩雑なキー操作を必要としない空気調和機が開示されている。(例えば、特許文献 1 参照)

20

【0004】

また、送信機を用いず、手叩き・音声認識等による運転、停止、設定変更等の音響指示を受け付けた時、機器側にてその指示入力を実行する旨の予告表示を出力して指示者に伝え、予告表示内容に対する指示者からの認可を意味する音響指示を再度受け付けた場合のみ動作・実行する空気調和機が開示されている。(例えば、特許文献 2 参照)

【0005】

一方、上記音声認識の弊害要因の 1 つであるノイズに対しては、帯域通過フィルタ (BPF: Band Pass Filter) などのフィルタを用いたり、定常雑音を音声データベースにも重畳させたりすることで音声認識の際の定常雑音の対策を行っている。

30

【0006】

【特許文献 1】特開平 03 - 255847 (図 1、第 3 頁左上欄 ~ 第 4 頁右上欄)

【特許文献 2】特開平 04 - 278141 (図 2、段落 0016)

【発明の開示】

【発明が解決しようとする課題】

【0007】

しかしながら、特許文献 1 に示される従来例のようにリモートコントローラ (以下、リモコンという) のボタンを押した後引き続き一定時間のみ単語認識する方式では、喋るタイミングを取るのが難しく、また、時間が限られているため、ユーザーは長い言葉は焦って早口になりがちであり、音声認識装置は追従できる速度と異なるため認識率が落ちるといった問題があった。

40

【0008】

また、特許文献 2 に示される従来例では、常に認識処理が動作しているため消費電力が大きく、また単語の始まりと終わりがどこなのか認識できないという問題があった。

【0009】

そこで、これらの問題を解決する方法として、予め定めた閾値を超えた音量のみ単語として認識する方法が考えられるが、スパイクノイズ (相対的に短く大きい音のノイズ: 瞬間音など) に弱く、ノイズが重畳した場合に大きく目的の単語から離れた特徴量が算出され音声認識部は違う単語を選択してしまうという問題が発生する。また、定常ノイズが大

50

きい場合の対策として周囲音に対して閾値を上げる方法が考えられるが、全ての設備機器が同じ環境にあるとは限らず閾値の最適値を見出せない。また定常雑音と音声と同じ音量の場合には音声と雑音の識別が不可能なため認識を開始できないという課題が発生する。さらにノイズ対策を行う場合、定常音が音声かを判別する必要があり、誤って判別した場合には必ず誤作動を伴い、回避できないという課題があった。

【0010】

本発明は上記の課題を解決するために為されたものであり、音声信号に背景ノイズやスパイクノイズなどのノイズが混入しても、自動的にあるいは簡単な方法で音声の開始時期を正確に検出してより正確な単語を選択できる省電力型のリモコン並びに家電機器を提供することを目的とする。

【課題を解決するための手段】

【0011】

この発明に係るリモコンは、周期的な所定期間を示す区間を複数有する複数区間毎にこの複数区間内の音声を取り込み、FFTにより周波数成分に変換し、所定の周波数帯における各周波数成分のレベルの積算量を一つ前の区間のそれと比較して差分を算出し、この差分が音声認識開始用の閾値を超えた第1の開始区間及び次に前記差分が音声認識終了用の閾値を超えた第1の終了区間を求め、さらに前記各区間毎にこの区間内の音声を取り込み、FFTにより周波数成分に変換し、所定の周波数帯における各周波数成分のレベルの積算量を一つ前の区間のそれと比較して差分を算出し、この差分が前記音声認識開始用の閾値を超えた第2の開始区間から次に前記差分が前記音声認識終了用の閾値を超えた第2の終了区間を求め、前記第1の開始区間と前記第2の開始区間に基づき開始区間を推定し、前記第1の終了区間及び前記第2の終了区間に基づき終了区間を推定し、推定された開始区間から推定された終了区間までの一連の複数区間の音声信号を出力する音響分析部と、音響分析部が出力した一連の複数区間の音声信号レベルを基に音声認識アルゴリズムを適用して前記複数区間の各区間毎に音声特徴ベクトルの抽出や他の補足計算を行う特徴量・補足計算部と、特徴量・補足計算部によって抽出された特徴ベクトルと予め保有するデータベースの語句とに基づいて音声認識を行う音声認識部と、音声認識部が認識した結果を操作指令として外部機器に送信する通信部とを備えたものである。

【発明の効果】

【0012】

本発明によれば、リモコンは、区間毎に音声信号を周波数に変換し、所定の周波数帯の各周波数成分のレベルの積算量を一つ前の周期のそれと比較して差分を算出し、これを閾値と比較することにより、音声の大きさの変化が所定値以上の音声が入力された時のみ音声認識を開始するので、簡便な方法で且つ正確に音声認識を実行することが可能である。また、複数区間毎に音声信号を周波数に変換し、所定の周波数帯の各周波数成分のレベルの積算量を一つ前の周期のそれと比較して差分を算出し、これを閾値と比較することにより、大まかな音声開始周期が分かるので、この周期毎に調べる方法と上記の区間毎に調べる方法とを組み合わせることで、精度を損なわず、より効率的に音声入力区間を検出することができる。

【発明を実施するための最良の形態】

【0013】

実施の形態1.

図1は、本発明の実施の形態1におけるリモコンのハードウェア構成を示す図であり、リモコン1は、図1に示すように、ユーザーの音声₁を音声信号の形で取り込む音声取込部2と、取り込んだ音声信号を増幅する増幅部3と、増幅された音声信号をMPU(Micro Processor Unit)が処理できるようにアナログ信号からデジタル信号に変換あるいは、MPUが出力したデジタル信号をアナログ信号である音声に変換するためのA/D・D/Aコンバータ4と、各種演算および処理を実行するMPU5と、音声信号を音声として出力する音声出力部6と、処理中あるいは処理済みの各種データ類を記憶するRAM7と、ソフトウェア及び音声データベース8を記憶するROM9と、エアコンに対して操作指

10

20

30

40

50

令信号を赤外線により送信する通信部10とから構成されている。また、A/D・D/Aコンバータ4、MPU5、RAM7、ROM9及び通信部10は共通のバスに接続されており、音声取込部2と音声出力部6は増幅部3を介してA/D・D/Aコンバータ4に接続されている。なお、エアコン11は通信部12を標準で装備している。

【0014】

また、図2は図1をMPU5の機能を複数の機能ブロックに分割して書き直した構成図であり、図中、図1と同符号は同一または相当部分を示す。また、音響分析部13と特徴量・補足計算部14と音声認識部15の各機能は、それぞれに対応する専用ソフトウェアをMPU5が実行することで実現される。なお、これらのソフトウェアはROM9に格納されており、必要時にRAM7に読み出された上でMPU5によって実行されるものである。

10

また、図9に示すように、リモコン1にはボタン16とONを示すLED(Light Emitting Diode)17とOFFを示すLEDが搭載されている。

【0015】

次に、実施の形態1の動作を説明する。例えば、電車の中では携帯電話の音が室内での声よりも大きくなるように、人間は通常より大きな雑音がある環境下ではより大きな音声で話すため、取り込んだ音響成分の前後区間では必ず差分が発生する。そこで、リモコン1のMPU5は、定期的に音声入力があるか否かを監視し、それ以外はスリープ状態に戻るだけの省電力モードを維持し、音声の大きさの変化が或る一定以上の音声が入力された時のみ音声認識を開始する。音声認識の際に、音響分析部13はまず、図3(a)に示すように周期的な一定期間を示す区間毎にこの区間内の音声を周期的に取り込み解析する。図3(a)の例では先ずA区間の音声を取り込み解析し、次にB区間の音声を取り込み解析し、次にC区間の音声を取り込み解析する。なお、この場合の一定区間(以後、単に区間という)とは、一般的な音声区間(20~40m秒)あるいは、これの整数倍長さの時間をいう。解析の際には、上記一定区間内の音声信号の振幅値または電力値(以下、音声レベルという)を11kHz(約100μ秒刻み)のサンプリングタイムで積算し、この積算量を一つ前の区間のそれと比較し、図3(b)に示すように差分が音声認識開始用の閾値を超えた区間Cを音声認識開始区間とし、そこから一連の複数区間の音声信号を音声認識処理のために特徴量・補足計算部14へ送る。

20

【0016】

この場合、音響分析部13は、予め音声認識開始の閾値と音声認識終了の閾値を設定しておき、図4(a)、(b)に示すように、上記周期的な所定期間である区間における音声レベルの積算量を一つ前の区間のそれと比較し、差分が音声認識開始の閾値を超えた区間を音声開始区間とし、予め決定した基準に従って、単語発生開始時点を決める。ここで、基準とは「音声開始用フラグを立てた区間の一つ前の区間も音声開始時点に含める」ということである。例えば、図4(a)に示すように差分前の音声レベルが大きくなった後の区間Cで終了しているのでこの区間Cを音声開始区間とする。また、差分発生後の区間Cの一つ前の区間Bにも単語の先頭が含まれている可能性もあるのでその区間Bやその区間内の音声波形の変曲点の現れる時点(100μ秒刻みで音声レベルの差分の変化が最大の時間的位置)を単語発生開始時点とする。

30

40

【0017】

また、音声終了区間も上記と同様の方法で検出可能である。音響分析部13は、図5(a)、(b)に示すように、上記区間における音声レベルの積算量を一つ前の区間のそれと比較し、差分が音声認識終了の閾値を超えた区間を音声終了区間とする。例えば、図5(a)に示すように差分前の音声レベルが小さくなる前の区間Eで終了しているが、差分発生後の次の区間Fにも単語の末尾が含まれている可能性もあるのでその区間Fやその区間内の音声波形の変曲点の位置を単語発生終了時点とし、単語発生開始から終了までの一連の音声区間の音声信号を特徴量・補足計算部14に送る。特徴量・補足計算部14は、音声認識アルゴリズムに応じて音声区間の特徴ベクトル抽出や他の補足計算を行い、得られた特徴ベクトルを音声認識部15に送る。そして、音声認識部15は特徴量・補足計算

50

部 1 4 から送られた特徴ベクトルと音声データベース 8 の語句とに基づいて認識処理を実行し、認識結果を指令として通信部 1 0 を介して空調機 1 1 などの家電機器に送りこの機器を操作する。

【 0 0 1 8 】

以上の通り、この実施の形態 1 によれば、前後の音声区間における音声レベル積算量同士の差分の値が閾値を超えた時を音声認識開始時点、音声認識終了時点とし、単語発声開始から終了までの一連の音声区間の音声信号を認識するので、音声入力区間が正確に検出でき、これにより音声認識の精度が高くなる。

【 0 0 1 9 】

実施の形態 2 .

実施の形態 1 では、時間領域での音声信号レベル積算量の差分により音声入力区間を検出したが、これに限らず、周波数領域で音声入力区間を検出することも選択可能である。この実施の形態 2 では、この周波数領域で音声入力区間を検出する方法について説明する。

音響分析部 1 3 は、周期的に上記一定区間の音声を取り込み、取り込んだ音声を高速フーリエ変換 (F F T : Fast Fourier Transform) により周波数成分に変換し、変換された周波数成分のうち、帯域通過フィルタ (B P F : Band Pass Filter) や低域通過フィルタ (L P F : Low Pass Filter) を用いて人間の音声周波数帯域 (相対的に低域のホルマント領域) のみの周波数成分について各周波数成分のレベルを 1 つの区間に渡って約 1 0 0 μ 秒 (1 1 k H z) 単位で積算した積算量を 1 つ前の区間のそれと比較して差分を算出し、この差分を F F T 逆変換したものと音声認識開始用の閾値とを比較する。そして、差分を F F T 逆変換したものが音声認識開始用の閾値を超えた区間から一連の複数区間を特徴量・補足計算部 1 4 に送る。特徴量・補足計算部 1 4 は、音声認識アルゴリズムに応じて音声区間の特徴ベクトル抽出や他の補足計算を行い、得られた特徴ベクトルを音声認識部 1 5 に送る。そして、音声認識部 1 5 は特徴量・補足計算部 1 4 から送られた特徴ベクトルと音声データベース 8 の語句とに基づいて認識処理を実行し、認識結果を操作指令として通信部 1 0 を介して空調機 1 1 などの家電機器に送りこの機器を操作する。

【 0 0 2 0 】

この方法では、人間の音声の周波数帯域 (所定の値よりも低い周波数帯) のみについて前後の区間における周波数成分の積算量の比較を行えるのでより正確な音声入力区間を検知することが可能になる。

【 0 0 2 1 】

一連の複数区間における音声開始区間の検出方法について説明する。音響分析部 1 3 は上記一定区間毎に F F T により音声信号を周波数成分に変換し、各周波数成分を 1 1 k H z のサンプリングクロックで積算した積算量を一つ前の区間のそれと比較し、各周波数成分毎に差分を計算し、得られた周波数成分毎の差分を F F T 逆変換したものが音声認識開始の閾値を超えた区間を音声開始区間とし、予め決定した基準に従って、単語発生開始時点を決する。例えば、図 4 (a) に示すように差分前の音声レベルが大きくなった後の区間 C で終了しているのこの区間 C を音声開始区間とする。また、差分発生後の区間 C の一つ前の区間 B にも単語の先頭が含まれている可能性もあるのでその区間 B やその区間内の音声波形の変曲点の現れる時点 (1 0 0 μ 秒刻みで音声レベルの差分の変化が最大の時間的位置) を単語発生開始時点とする。

【 0 0 2 2 】

また、一連の複数区間における音声終了区間も検出可能である。音声区間終了時にも同様の方法で、音響分析部 1 3 は音声認識開始検知後もある一定区間ごとに F F T により周波数成分に変換し、周波数成分の積算量を時間領域の一つ前の区間のそれと比較し、各周波数成分毎に差分を計算し、得られた周波数成分毎の差分を F F T 逆変換したものが音声認識終了の閾値を超えた区間を音声終了区間とし、予め決定した基準に従って、単語発生終了時点を決する。例えば、図 5 (a) に示すように差分前の音声レベルが小さくなる前の区間 E で終了しているが、差分発生後の次の区間 F にも単語の末尾が含まれている可

10

20

30

40

50

能性もあるのでその区間Fやその区間内のある部分などを単語発声終了とし、単語発声開始から終了までの一連の音声区間の音声信号を特徴量・補足計算部14に送る。特徴量・補足計算部14は、音声認識アルゴリズムに応じて音声区間の特徴ベクトル抽出や他の補足計算を行い、特徴ベクトルを音声認識部15に送る。そして、音声認識部15は特徴量・補足計算部14から送られた特徴ベクトルに対して音声データベース8の語句を適用して認識し、その結果を基に機器を操作する。

【0023】

また、本実施の形態2のように周波数領域で音声区間を検出する方法では、LPFやBPFなどのフィルタを用いて、人間の音声周波数帯のみに絞り、この人間の音声周波数帯内で各周波数成分のレベルを時間領域で各区間毎に積算した値を比較するので、定常ノイズなど人体以外の音(ノイズ)に対して強く、また、単語発生区間に含まれる無音にも強い。前記音響分析部13は変化があった周波数帯域の周波数成分のみ、またはその周波数帯の周波数成分の重みを重くしてこれらの周波数成分を再構成し、これらの周波数成分をFFT逆変換した結果を音声信号として出力し、この音声信号を基に音声認識を行うためノイズの影響はより少なくなる。また、予め特徴量算出に用いる信号レベルの範囲(上限値あるいは下限値あるいはその両方)を決め、その範囲以外の信号レベルを持つ信号を特徴量算出の対象から排除してもよい。例えば、図6(b)に示すように音声周波数帯に信号レベルが予め設定した上限値を超える周波数成分(スパイクノイズ)が重畳した場合には、このスパイクノイズは特徴量算出に使用しないので、このノイズによる影響を受けな

10

20

【0024】

なお、上記の例では周波数成分の信号レベルが上限値を超える周波数成分(スパイクノイズ)が重畳した場合には特徴量算出に使用しないようにしたが、これに限らず、周波数成分の信号レベルが上限値を超える周波数成分(スパイクノイズ)の信号レベルの重みを軽くするようにしても良い。また、上限値を超える周波数成分(スパイクノイズ)の代わりに隣接する周波数成分のレベルの平均値を計算して、その値に強制的に置き換えてもよい。いずれの場合も上記と同様の効果を奏する。

【0025】

また、音声認識部15のアルゴリズムは、DPマッチングや隠れマルコフモデル(HMM: Hidden Markov Model)などがあるが、認識エンジンに合わせて音声取得と音声認識を並列処理できる場合には、音声認識開始を検出したら単語発生区間の終了を待たずに並列処理にて順次処理する。また、同一のMPU5で処理を行う場合、タスク処理などを用いて並列処理を行う。これにより、処理の高速化を図ることができる。

30

【0026】

さらに、上記HMMに基づき過去のフレームから得られた音声特徴量の変遷に沿った補正処理を行うことで、雑音の突発的な変化による影響を抑えることができる。

また、事前に登録した複数区間の音声レベルや定期的(周期的)に取り込んだ音声区間における音声レベルの解析結果を基に、音声認識の対象とするべき周波数帯や差分による音声区間検出の閾値を決定する事も可能である。この場合、雑音が多く、音声の差分が小さくなる場合でも対応可能である。音響分析部13は、差分がある一定範囲内で推移している時に定期的に取り込んだ音が定常雑音やスパイクノイズなどのノイズなのか音声なのかを上記の閾値を用いて判断する。これにより閾値を決定する要因である音声とノイズを判断でき、誤判定なく自動学習により閾値を変化させる事も可能である。この判定でも遠くの人間の声が混じっていたりして完全ではないので、判定は複数回繰り返した結果と比較し、ノイズのデータベースと比較したり、複数回の音量レベルを基に例えば平均値を計算したりして定常雑音の大きさを判断したりするので、たとえ瞬時的に誤判定が起こり閾値が変化しても正常に収束可能である。

40

【0027】

図7はこの実施の形態1及び2におけるリモコン1のMPU5の音声認識の処理を示すフローチャートである。次に、MPU5の音声認識の動作について図7を用いて説明する

50

。MPU5は、ステップS1にてマイクなどの音声取込部2から増幅器及びA/D変換コントローラを介して音声取込を行う。次に、ステップS2において音響分析を行い、その結果をステップS4にてRAMに保存する。ステップS2の音響分析は具体的には周期的(所定期間)に取り込んだ音声を11kHzのサンプリングクロックで各区間毎に積算する。FFTを使用する方法では、各音声帯周波数成分の音声レベル(音声信号の振幅または電力)を積算する。または、複数区間の音声周波数帯の音声レベルを積算する。ステップS3において一つ前の周期がない場合はもう一周期取り込む、一つ前の周期が存在する場合にはステップS5に進む。ステップS5では、一つ前の周期で同様に算出された積算値との差分を算出し、図4のように音声開始の閾値以上の差分がある場合は音声開始とみなし、その区間を音声開始区間とする。さらに、差分が音声区間終了の閾値よりも低い場合にはステップS1~S5を繰り返し実行して同様に音声を区間ごとに差分をとる動作を継続する。

10

【0028】

そして、差分が図5のように音声区間終了の閾値以上の差分がある場合はその区間を音声終了区間とする。ステップS6にて音声終了区間を検出後、音声認識を開始する。ステップS7にて、音声認識アルゴリズムの種類に応じて音声区間の特徴量や他の補足計算を行う。例えば音声認識にHMMを用いる場合では入力音声からメルスペクトラム分析により音響特徴量としてメルケプストラム、このメルケプストラムの動的特徴量(メルケプストラムの時間領域での変化量)、logパワーの動的特徴量などの特徴ベクトルを抽出する。次に、ステップS8にてこの特徴ベクトルに対してHMMによる認識を行う。この場合、認識したい対象を一つのHMMとして表現し、そのHMMが、実際に取得された音声の特徴ベクトル列を生成する確率を計算し、最も高い確率で出力するHMMを認識結果とする。その結果をステップS9にてリモコン1が保持する音声出力部6に出力する。認識結果の確認がない場合は、そのまま通信部10を介して、エアコン11などの家電機器に操作信号を送る。

20

【0029】

上記に説明したように、音響分析部13は、まず、図3のように或る一定区間の音声を周期的に取り込み解析する。具体的には、音響分析部13は、或る一定区間の音声レベルの積算量またはFFTにより変換した周波数成分を時間領域における一つ前の区間のそれと比較し、差分または各周波数成分の差分をFFT逆変換により変換した積算量が音声認識開始の閾値を超えた区間から一連の音声区間の音声信号を特徴量・補足計算部14に送る。特徴量・補足計算部14は、音声認識アルゴリズムに応じて音声区間の特徴ベクトル抽出や他の補足計算を行い、特徴ベクトルを音声認識部15に送る。そして、音声認識部15は特徴量・補足計算部14から送られた特徴ベクトルに対して音声データベース8の語句を適用して認識し、その結果を基に機器を操作する。

30

これにより、環境に合わせて音量、周波数の少なくとも一方を使用することで、より精度の高い単語発生検知を行うことができる。そして少なくとも一方が閾値を超えた場合に音声認識を開始する(どちらも閾値を超えないと音声認識を開始しない)ようにしておけば、予め決められた数値などで判断した場合と異なり、個人差を補いきれず、認識を開始しない(勝手に認識を開始する)といった問題を低減することが可能である。

40

【0030】

なお、以上の例では、各区間毎に当該区間内の音声信号からFFTにより得られた各周波数成分のレベルの積算量を一つ前の区間のそれと比較して差分を算出し、この差分が閾値を超えた区間を音声入力開始区間または音声入力終了区間であると判断したが、これに限らず、複数区間毎に当該複数区間(以下、周期という)内の音声信号からFFTにより得られた各周波数成分のレベルの積算量を一つ前の周期のそれと比較して差分を算出し、これを閾値と比較するようにしてもよい。これにより、大まかな開始周期が分かるので、この周期毎に調べる方法と上記の区間毎に調べる方法とを組み合わせることで、精度を損なわず、より効率的に音声入力区間を検出することができる。

【0031】

50

以上のように、この実施の形態 2 によれば、人間の音声の周波数帯域（所定の値よりも低い周波数帯）のみについて前後の区間における音声レベルの積算量の比較を行い、それ以外の定常ノイズを排除し、また、スパイクノイズのような瞬時的な所定値よりも大きなノイズも低減するので、より正確な音声入力区間を検知することが可能になり、これにより音声認識の精度がさらに高くなる。

【 0 0 3 2 】

実施の形態 3 .

実施の形態 1 及び実施の形態 2 では音声レベルの積算量の差分に基づいて音声入力開始時点を特定したが、より簡単な方法として図 9 に示すボタン 1 6 をユーザーが押下操作することで音声入力開始時点をリモコンに通知させるようにしてもよい。この実施の形態 3 10
では、このような実施の形態について説明する。

ユーザーが図 9 のボタン 1 6 を押した時にボタン 1 6 は押下信号を発生する。M P U 5 はボタン 1 6 からの押下信号を図示しない入出力部を介して受信すると、音声認識を開始する。そして、ユーザーがボタンを押し続けている間はボタン 1 6 から押下信号が継続して発生するので、M P U 5 はこの押下信号を受信している間はユーザーの音声を音声取り込み部 2 に取り込ませ、音声認識を行う。従って、ユーザーはボタンを押し続けている間自由に話すことができるので焦って話す必要はない。

【 0 0 3 3 】

図 8 はこの実施の形態 3 におけるリモコン 1 の M P U 5 の音声認識の処理を示すフローチャートである。次に、M P U 5 の音声認識の動作を図 8 を用いて説明する。 20

M P U 5 は、ステップ S 8 1 において、ボタン付きか否かを調べ、ボタン付きである場合には、ボタンが押されるまで S 8 2 を繰り返し実行して待つ。ステップ S 8 2 において M P U 5 はユーザーによるボタンの押下信号を検出すると、音声取り込みを開始する（ステップ S 1 ）。ステップ S 1 ~ S 9 の動作は図 7 と同じなので説明を省略する。ステップ S 8 3 において、M P U 5 はボタンが押下されなくなったか否かを調べ、まだ押されていればステップ S 1 に戻り、S 1 ~ S 9 の音声認識処理を再開する。ステップ S 8 3 において、ボタンが押されなくなったことを検出したら、音声認識処理を終了する。また、S 8 1 において、ボタンがない場合の S 1 ~ S 9 の動作は図 7 と同じなので説明を省略する。

【 0 0 3 4 】

以上のように本実施の形態 3 によれば、ユーザーはボタンを押ししている間自由に話すことができるので焦って話す必要はない。また、リモートコントローラは信号を受信した後に、L E D などの表示装置や音声応答などで認識結果をユーザーに伝え、ボタンを用いて音声認識により実行およびキャンセル処理を行うことができ、結果を選択できる。さらにその結果を学習でき、音声認識データベースに反映できるのでより正確な認識を行うことができる。 30

【 0 0 3 5 】

音声認識の認識率がおちる場合には音声認識完了後、受け付けた内容を L E D で表示したり、音声で伝えたりユーザーに受け付けた内容を表示する機能を選択できる。ユーザーはその内容で良ければボタンを押すなど設定したアクションを行い、実行・中止を選択できる。例えばコントローラのインターフェースとしては音声取得部分のマイク 3、ボタン 1 つ、L E D 2 つのコントローラでは音声認識後、認識した内容を L E D に点灯して表示する。そしてユーザーは決定ならボタンを一回押して決定する。L E D は処理を実行する際に L E D を点滅させる。キャンセルなら L E D が消灯するまでボタンを押し続ける。という様に小部品で色々な操作が可能になる。 40

【 0 0 3 6 】

なお、マイク、音声出力機器、ボタンなどは他の機器で代用することが可能である。例えば、リモコンにマイクのみを装着し、音声を入力したらネットワークを通じてパソコンや H D D - D V D レコーダーなどに転送し、音声認識処理を代わって行わせる。つまり、音声認識の機能をパソコンなどの高速なマイコンに代わって処理してもらい、認識結果を受け取ることによりリモコン自体は低能力のマイコンで実現できる。また、認識結果の確 50

認を行わせる場合、テレビや冷蔵庫LEDなど他の機器で表示させる事も可能である。もちろん、機器の情報をリモコンで表示し、機器の衣装デザインの自由度を高める事もできる。他の機器と連動させる場合、それぞれの機器に専用のS/Wをインストールしておくか、各機器のOSの他に仮想システムを共通で立ち上げておき、割り込みで割り込ませるなどする。個々のインターフェースなどを複数の機器に行わせるのではなく、S/Wをサーバなどから提供し、携帯電話など1つの機器で完全に全ての機能をコントロールさせることができる。

【0037】

本発明のリモートコントローラは複数の機器で操作する場合や違う機器のリモコンとして利用する場合にIrDAなどのネットワーク経由やUSB、SDカード(商標)などを用いてS/Wや音声認識などのデータベースを書き換えたり、H/Wにより付け替えたりすることで対応可能である。例えば無線通信方法を変更する場合、CFタイプ(商標)の特定省電力無線の無線カードをZigBeeの無線カードに付け替えるだけで違う無線方式の機器を操作できる。またminiSDカード(商標)などのIDコードやコマンドが入ったカードを入れ替えるだけで違う機器や後継機種を操作可能になる。パソコンや携帯電話でminiSDカードにデータを入れてリモートコントローラに差し込むだけで最新のデータや機器が操作可能になる。これらの記憶装置を使って操作する機器情報を記録できる。例えば、エアコンなどでは故障や定期メンテナンスなどを行うが、無線タグとそのリーダ・ライターをリモコンに具備しておき、機器の運転状況やライフサイクルを確認する為の無線タグなどの記憶媒体を利用する場合、音声操作などによって音声認識での操作結果をリモコンのリーダ・ライターにて無線タグに書き込ませる事ができる。

【0038】

双方向通信可能な場合は電流値、故障情報、サーミスタ温度などの情報を受けて書き込むことができる。サービスマンはエアコンに接続して情報を抜き出すのではなく、リモコン位置で作業が可能になり、より作業難度の敷居が下がる。またユーザーは故障した場合に無線タグやminiSDカードなどのメディアを送っても良いし、そこから取り出した情報をメールでサービスセンターに送信してもよい。これによりサービスマンは故障状態をより把握でき、修理時に持っていくサービス部品も限定できる。環境情報、運転状況により起こる不具合解析もサーミスタ検出温度などの情報もあれば容易になる。他の家電機器からの湿度情報などと複合的に判断可能である。またそれらの情報を自動的にリモコンの記憶媒体に書き込むことも可能である。勿論、リモコン自体のFlashメモリーに書き込む事も可能である。

【0039】

また、押している間のみ音声を検出するモードもある。音声区間が一定では自由な発音ができない。押している間のみ音声入力であれば自分のタイミングで話す事ができ、ボタンを押した後に慌ててしゃべる必要もなく、通常会話と同じ感覚で使用することが可能になる。リモートコントローラの放置状態が悪く、ボタンが入力状態になり続けいても図2のように差分のみを定期的に計測するので音声認識処理を行わないので消費電力を抑えられる。また、ある一定時間以上ボタンが押し続けられるとボタンが一度OFFするまで異常と判断しスリープモードに入ることも可能である。

【0040】

また、音声認識精度が低い場合、音声により再度決定の意の発声をさせたり、認識結果を確認させたりするが結果的に受理できなかつたりする。本発明のリモートコントローラでは、音声により階層の深い操作を選択し、ボタン操作1つで実行または中止を行う。例えば図9のリモコン1では、ボタンを押している間、音声認識を行い結果をLEDで表示する。ONなら左のLED17点灯で表示される。決定の場合はボタン16を一回押すと左のLED17が点灯状態になるので決定を受け付けたことを示す。中止の場合はボタン16を押し続けると左右のLED17、18が交互に点滅したら中止を受け付けたのでボタン16を放す。温度UPなら左のLED17が点滅する、というような素子数の少ない構成でリモコンを制作可能である。双方向通信で音声認識処理を他の機器に行わせる場合

10

20

30

40

50

、音声を送信し、結果だけを受け取りLEDで表示することも可能でありより構成部品の少ない構成になる。

【0041】

なお、ボタン16が静電センサでも実現可能であり、手に取るだけで音声認識が開始されるのでボタンを意識する必要はない。この場合、音声で依頼した処理の音声認識結果を確認し、誤認識で処理を中止する場合、上述したように同じリモコンを握り続け（ボタンを押し続け）ることも可能であり、指をスライドさせると中止するなど様々な対応が可能である。静電センサではボタンが必要なく、基板パターンで実現できるのでコストも抑えられる。使い勝手向上のためにボタン部分に凹凸をつけたり入力時に音を出力したり、誤動作防止のために同時にボタンが押された場合は反応しない、また手にとって音声認識後、もう1つ押さないと実行させないなどの方法が選択できる。

10

【0042】

また、指紋センサを設けて個人識別を行い、音声認識のアルゴリズムやデータベースを変更することにより、より認識精度があがる。前述したように他の機器に機能を移行できるので指紋センサとマイクのデータを転送したり、他の機器からの個人識別情報を用い、音声認識のアルゴリズムやデータベースを変更することも可能である。個人識別を取得する方法としては、家電機器、例としてパソコンや携帯電話などの指紋センサ、ホームセキュリティのIDチェックの情報やカメラの画像認識結果、他の製品につけられた無線タグのリーダー・ライターなどを利用して行うことが可能である。

20

【0043】

本発明のリモートコントローラから以上の機能を携帯電話にS/Wをダウンロードし使用できる。また機器本体に組込んだり、それぞれの機能を分散して処理したり、他の機器からの情報を積極的に利用したりできる。また、応用の一例として、エアコンのリモコンを例にとると、パソコンの指紋認証の情報を得て個人を識別し、認識率を上げ、認識結果を個人の状態に合わせて制御する。「暑い」の発声で暑がりの人は室温を下げると同時に気流制御で風を当てるなどして快適な生活環境を実現できる。

【0044】

本発明のリモコンは家電製品に内蔵することもでき、家電製品が近くにある場合やマイクの感度が良好で音声認識モジュールの精度が高い場合は有効である。

【0045】

また、本リモコンは音声認識だったが、画像認識でも同様にCMOSやCCDなどの画像センサを用いてジェスチャーや指の動き、顔認識による環境設定を行い快適な生活環境を実現できる。

30

【図面の簡単な説明】

【0046】

【図1】本発明の実施の形態1におけるリモコンのハードウェア構成を示す図である。

【図2】図1をMPUの機能を複数の機能ブロックに分割して書き直した構成図である。

【図3】音声区間検出方法を示す図である。

【図4】音声区間検出方法を示す図である。

【図5】音声区間検出方法を示す図である。

40

【図6】音声区間検出方法を示す図である。

【図7】本発明の実施の形態1、2におけるリモコンの音声認識フローチャートである。

【図8】本発明の実施の形態3におけるリモコンの音声認識フローチャートである。

【図9】本発明を利用したエアコンコントローラの一例を示す図である。

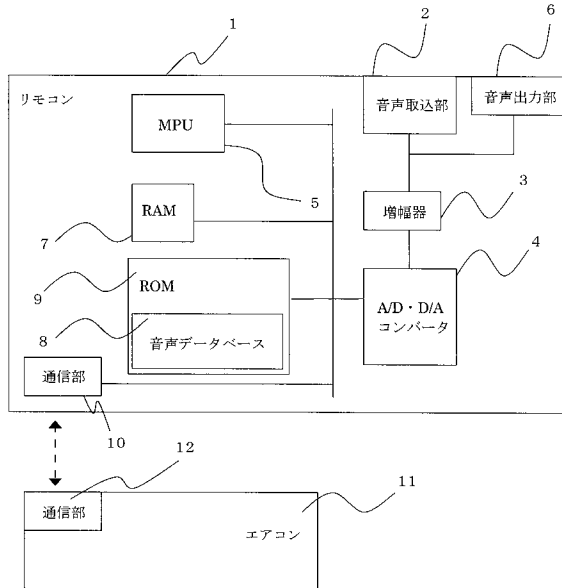
【符号の説明】

【0047】

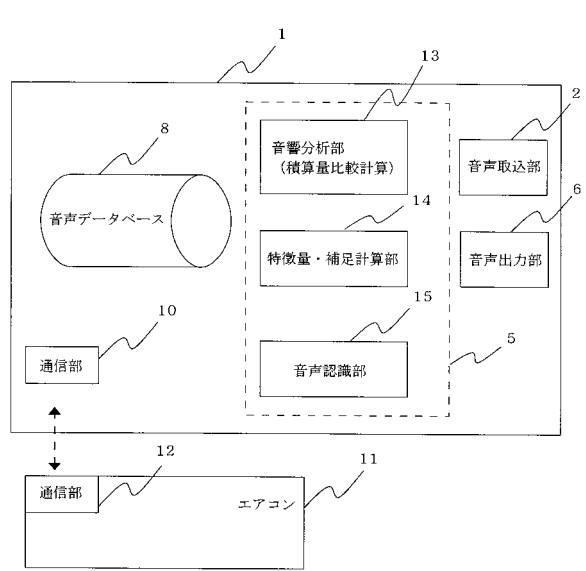
1 リモコン、2 音声取込部、3 増幅器、4 A/D・D/Aコンバータ、5 MPU、6 音声出力部、7 RAM、8 音声データベース、9 ROM、10 通信部、11 エアコン、12 通信部、13 音響分析部、14 特徴量・補足計算部、15 音声認識部、16 ボタン、17 LED、18 LED。

50

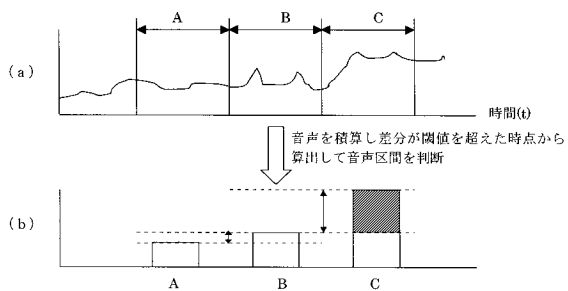
【図1】



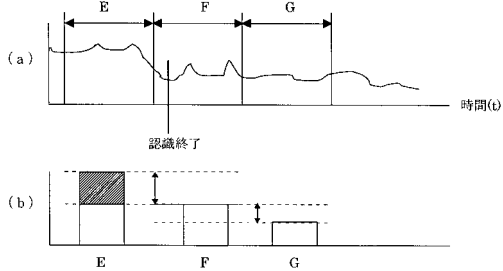
【図2】



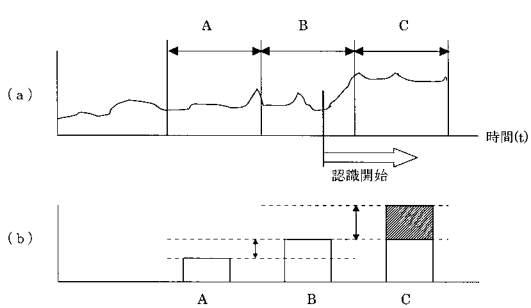
【図3】



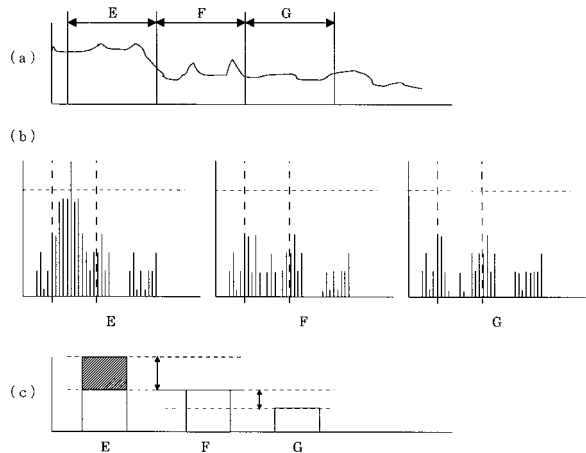
【図5】



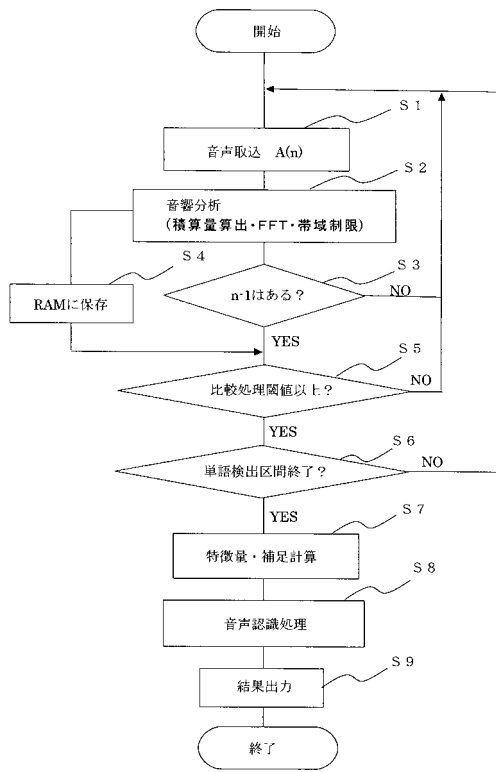
【図4】



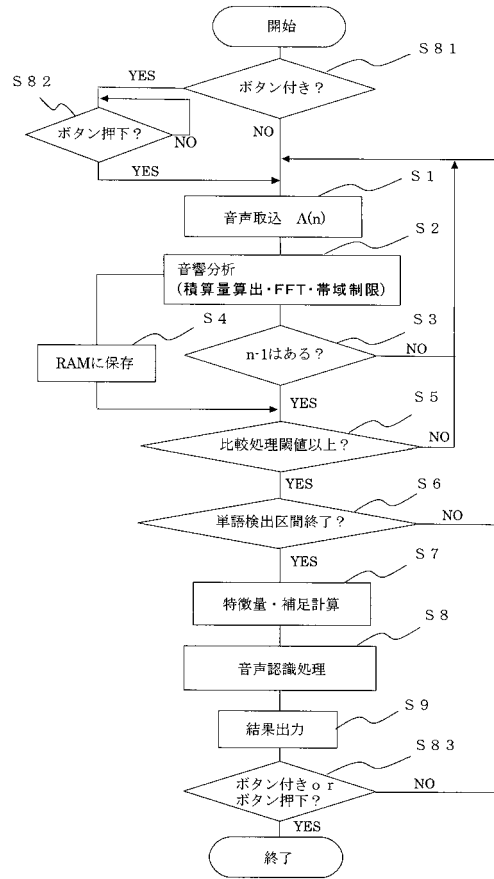
【図6】



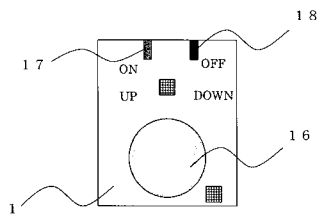
【図7】



【図8】



【図9】



フロントページの続き

- (72)発明者 矢野 裕信
東京都千代田区丸の内二丁目7番3号 三菱電機株式会社内
- (72)発明者 望月 昌二
東京都千代田区丸の内二丁目7番3号 三菱電機株式会社内

審査官 前田 祐希

- (56)参考文献 特開平03-253899(JP,A)
特開平06-230799(JP,A)
特開平04-178698(JP,A)
特開昭62-129899(JP,A)
特開昭56-135898(JP,A)
特開2001-265367(JP,A)
特開2004-294814(JP,A)
特開2000-163098(JP,A)
特開2004-294659(JP,A)

- (58)調査した分野(Int.Cl., DB名)
G10L 15/00 - 17/00