(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第3726448号

(P3726448)

(45)発行日　平成17年12月14日(2005.12.14)　(24)登録日　平成17年10月7日(2005.10.7)

(51)Int.Cl.[7]	F I		
G10L 15/28	G10L	3/00	571K
G10L 11/02	G10L	3/00	571A
G10L 15/04	G10L	3/00	513A

請求項の数 16　(全 16 頁)

(21)出願番号　特願平9-280670
(22)出願日　平成9年10月14日(1997.10.14)
(65)公開番号　特開平10-312194
(43)公開日　平成10年11月24日(1998.11.24)
審査請求日　平成15年1月16日(2003.1.16)
(31)優先権主張番号　特願平9-57856
(32)優先日　平成9年3月12日(1997.3.12)
(33)優先権主張国　日本国(JP)

前置審査

(73)特許権者　000002369
セイコーエプソン株式会社
東京都新宿区西新宿２丁目４番１号
(74)代理人　100095728
弁理士　上柳　雅誉
(74)代理人　100107076
弁理士　藤綱　英吉
(74)代理人　100107261
弁理士　須澤　修
(72)発明者　稲積　満広
長野県諏訪市大和３丁目３番５号　セイコーエプソン株式会社内
(72)発明者　相澤　直
長野県諏訪市大和３丁目３番５号　セイコーエプソン株式会社内

最終頁に続く

(54)【発明の名称】認識対象音声検出方法およびその装置

(57)【特許請求の範囲】

【請求項１】

音入力手段に入力された音声を認識して、前記認識の結果に対して何らかの動作を行う音声認識装置における認識対象音声検出方法であって、

前記音入力手段を間欠駆動し、当該間欠駆動される音入力手段が動作状態の間に入力された音に対し、当該入力された音が音声であるかを判定する処理を、複数段階の工程に分けて段階的に行い、現在処理中の工程での処理結果が当該処理中の工程に設定された条件を満たした以降に次の段階の工程が動作し、段階を経るに従って、消費電流が大きく、かつ、前記入力された音が音声であるかの判定確度の上がる処理に移行し、

それぞれの工程における処理において、その工程に設定された条件が満たされない場合は、それぞれの工程を非動作状態に戻し、

前記複数段階の工程における処理により、前記入力された音が音声であると判定された場合は、前記音入力手段の動作状態を維持し、前記入力された音の認識処理へ移行することを特徴とする認識対象音声検出方法。

【請求項２】

音入力手段に入力された音声を認識してその認識結果に対して何らかの動作を行う音声認識装置における認識対象音声検出方法であって、

前記音入力手段を間欠駆動し、

間欠駆動される音入力手段が動作状態の間に入力された音のレベルを検出し、

そのレベルの大きさから音の有無を判定し、音が無いと判定した場合は、非動作状態に戻

る第１の処理工程と、
前記第１の処理工程で音が有りと判定された以降に動作を開始し、前記入力された音が雑音であるか音声らしき音であるかを大まかに判定し、音声らしき音ではないと判定した場合は、非動作状態に戻る第２の処理工程と、
前記第２段階の処理工程で音声らしき音と判定された以降に動作を開始し、前記音声らしき音と判定された音が音声であるか否かを判定し、音声であると判断した場合は、前記音入力手段の動作状態を維持し、前記音声であると判断された音の音声特徴データを認識部側に渡し、音声でないと判断した場合には、非動作状態に戻る第３の処理工程と、
を有したことを特徴とする認識対象音声検出方法。
【請求項３】
前記第１の処理工程は、前記音入力手段が動作状態の間に入力された音の平均パワーを求め、当該求められた平均パワーと基準レベルとを比較して音の有無を判定し、音が無いと判定した場合は、非動作状態に戻ることを特徴とする請求項２記載の認識対象音声検出方法。
【請求項４】
前記第１の処理工程は、前記音入力手段が動作状態の間に入力された音を、人間の音声の周波数帯域を含む周波数帯域とそれ以外の周波数帯域に分けて少なくとも一方の周波数帯域の平均パワーを求め、当該求められた平均パワーの値を基に音を判定し、人間の音声ではないと判定した場合は、非動作状態に戻ることを特徴とする請求項２または３記載の認識対象音声検出方法。
【請求項５】
前記第２の処理工程は、前記第１の処理工程での設定された条件を満たした音信号に対し、当該音信号の継続時間を計測し、当該計測した継続時間を基に音声らしき音か否かを判定し、音声らしき音ではないと判定した場合は、非動作状態に戻ることを特徴とする請求項２～４のいずれかに記載の認識対象音声検出方法。
【請求項６】
前記第２の処理工程は、前記第１の処理工程での設定された条件を満たした音信号に対し、当該音信号の所定時間内における零交差数を計測し、当該計測した零交差数を基に音声らしき音か否かを判定し、音声らしき音ではないと判定した場合は、非動作状態に戻ることを特徴とする請求項２～５のいずれかに記載の認識対象音声検出方法。
【請求項７】
前記第３の処理工程は、前記第２の処理工程での設定条件を満たした音信号に対し、音声特徴抽出処理を行い、前記音声特徴抽出処理により抽出された音声特徴データを基に、入力音が音声であるか否かを判断して音声であると判断した場合に、前記音声であると判断した音の特徴データを認識部側に渡し、音声でないと判断した場合は、非動作状態に戻ることを特徴とする請求項２～６のいずれかに記載の認識対象音声検出方法。
【請求項８】
前記認識部では、設定されたキーワードを含む音声特徴データのみを認識処理することを特徴とする請求項７記載の認識対象音声検出方法。
【請求項９】
音入力手段に入力された音声を認識してその認識結果に対して何らかの動作を行う音声認識装置における認識対象音声検出装置であって、
前記音入力手段を間欠駆動する間欠駆動制御手段と、
前記間欠駆動制御手段により間欠駆動される音入力手段が動作状態の間に入力された音に対し、前記入力された音が音声であるかを判定する処理を、複数段階に分けて段階的に行うそれぞれの処理手段を有し、
現在処理中の処理手段による処理結果が、当該処理中の処理手段に設定された条件を満たした以降に、次の段階の処理手段が動作し、段階を経るに従って、消費電流が大きく、かつ、前記入力された音が音声であるかの判定確度の上がる処理に移行し、
それぞれの処理手段における処理において、その処理手段に設定された条件が満たされ

ない場合は、それぞれの処理手段を非動作状態に戻し、
前記複数段階の処理手段における処理により、前記入力された音が音声であると判定された場合は、前記音入力手段の動作状態を維持し、前記入力された音の認識処理へ移行するようになっていることを特徴とする認識対象音声検出装置。
【請求項１０】
音入力手段に入力された音声を認識してその認識結果に対して何らかの動作を行う音声認識装置における認識対象音声検出装置であって、
前記音入力手段を間欠駆動する間欠駆動制御手段と、
前記間欠駆動制御手段により間欠駆動される音入力手段が動作状態の間に入力された音のレベルを検出し、前記検出した音のレベルの大きさから音の有無を判定し、音が無いと判定した場合は、非動作状態に戻る入力レベル判定手段と、
前記入力レベル判定手段で音が有りと判定された以降に動作を開始し、前記入力された音が雑音であるか音声らしき音であるかを大まかに判定し、音声らしき音でないと判定した場合は、非動作状態に戻る音判定手段と、
前記音判定手段で前記入力された音が音声らしき音と判定された以降に動作を開始し、前記音声らしき音と判定された音が音声であるか否かを判定し、音声であると判定した場合は、前記音声であると判定された音の音声特徴データを認識部側に渡し、音声でないと判断した場合には、非動作状態に戻る音声判定手段と、
を有し、
前記音判定手段が音声であると判定した場合には、前記音入力手段の動作状態を維持するようになっていることを特徴とする認識対象音声検出装置。
【請求項１１】
前記入力レベル判定手段は、前記音入力手段が動作状態の間に入力された音の平均パワーを求め、当該求められた平均パワーと基準レベルとを比較して音の有無を判定し、音が無いと判定した場合は、非動作状態に戻ることを特徴とする請求項１０記載の認識対象音声検出装置。
【請求項１２】
前記入力レベル判定手段は、前記音入力手段が動作状態の間に入力された音を、人間の音声の周波数帯域を含む周波数帯域とそれ以外の周波数帯域に分けて少なくとも一方の周波数帯域の平均パワーを求め、当該求められた平均パワーの値を基に音を判定し、人間の音声ではないと判定した場合は、非動作状態に戻ることを特徴とする請求項１０または１１記載の認識対象音声検出装置。
【請求項１３】
前記音判定手段は、前記入力レベル判定手段での設定された条件を満たした音信号に対し、前記音信号の継続時間を計測し、前記計測した継続時間を基に音声らしき音か否かを判定し、音声らしき音ではないと判定した場合は、非動作状態に戻ることを特徴とする請求項１０～１２のいずれかに記載の認識対象音声検出装置。
【請求項１４】
前記音判定手段は、前記入力レベル判定手段での設定された条件を満たした音信号に対し、前記音信号の所定時間内における零交差数を計測し、前記計測した零交差数を基に音声らしき音か否かを判定し、音声らしき音ではないと判定した場合は、非動作状態に戻ることを特徴とする請求項１０～１３のいずれかに記載の認識対象音声検出装置。
【請求項１５】
前記音声判定手段は、前記音判定手段での設定された条件を満たした音信号に対し、音声特徴抽出処理を行い、前記音声特徴抽出処理により抽出された音声特徴データを基に入力音が音声であるか否かを判断して音声であると判断した場合に、前記音声であると判断した音の音声特徴データを認識部側に渡し、音声でないと判断した場合は、非動作状態に戻ることを特徴とする請求項１０～１４のいずれかに記載の認識対象音声検出装置。
【請求項１６】
前記認識部では、設定されたキーワードを含む音声特徴データのみを認識対象音声とし

て認識処理することを特徴とする請求項１５記載の認識対象音声検出装置。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】
本発明は、入力音声を認識してその認識結果に基づいて何らかの動作を行う音声認識装置において、装置側が入力音声を常に待つ状態となっている場合、入力音声を効率よく検出して消費電流を少なくする認識対象音声検出方法およびその装置に関する。
【０００２】
【従来の技術】
最近、音声認識を用いた機器が様々な分野で実用化されてきている。この種の機器は、スイッチを入れて初めて認識動作を開始することで十分な機能を果たすものもあるが、音声が入力されると、直ちに入力音声を認識してその認識結果に基づいた動作を行うというように、常に、入力音声を待ち状態としておく必要のあるものもある。
【０００３】
後者の例としては、たとえば、ユーザが時刻を問い合わせると現在時刻を応答する時計などがある。この種の機器は、乾電池で動作するものが殆どであり、また、機器の小型、軽量化を考えたとき、乾電池は小容量のものを使用することが望ましく、かつ、長時間、電池交換をしないで済むことが望まれる。
【０００４】
しかし、この種の機器は、常に音声入力を待ち状態にしておく必要があるため、待ち状態においても、常に電流が消費されることになり、その消費電流を如何に小さく抑えるかが大きな課題である。
【０００５】
音声を常に待つ状態としておくには、マイクロホンやアンプなどの音声検出回路を常に動作可能状態としておく必要がある。この種の機器に一般的に用いられているコンデンサマイクロホンの消費電流は、５００μＡ程度であり、マイクロホンで入力した音声信号を処理するアンプの消費電流も同様に５００μＡ程度である。
【０００６】
したがって、これらの音声検出回路の消費電流は１ｍＡ程度となる。この状態を保持するとすれば、一年間に8.76ＡＨｒ．を消費することになる。この値は、単１のアルカリ電池に相当するものであり、安価な単１マンガン電池の２本分に相当する。
【０００７】
機器の小型化、軽量化さらには価格を考えたとき、電池は単３以下を用いるのが望ましいが、上述した例では、電池の寿命はごく短いものとなってしまうという問題があった。
【０００８】
これに対処するには、必要なときだけスイッチをオンして音声入力を可能とすることも考えられるが、動作を行わせるためにその都度スイッチを入れるのでは、この種の音声認識を用いた装置としてのメリットが全く失われることになり、現実的な方法ではない。また、他の方法として、特公昭６１－５４１９１がある。この従来技術は、アラーム付き電子時計において、アラームセット時刻が到達した以降のアラーム装置の動作を、音声入力によって制御可能としたものである。
【０００９】
【発明が解決しようとする課題】
しかしながら、前述の特公昭６１－５４１９１で示される技術は、アラーム信号などで音声入力のタイミングを制御するものであり、任意のタイミングで音声入力を行うことができないので、使い勝手が悪い問題があった。
【００１０】
そこで、本発明は、入力音声を効率よく検出することで、認識対象音声の入力待ち状態における機器の消費電流を小さく抑え、乾電池を使用する機器においては、小容量の電池で長時間使用することができる認識対象音声検出方法およびその装置を実現することを目的

としている。

【００１１】

【課題を解決するための手段】

本発明の認識対象音声検出方法は、音入力手段に入力された音声を認識して、前記認識の結果に対して何らかの動作を行う音声認識装置における認識対象音声検出方法であって、前記音入力手段を間欠駆動し、当該間欠駆動される音入力手段が動作状態の間に入力された音に対し、当該入力された音が音声であるかを判定する処理を、複数段階の工程に分けて段階的に行い、現在処理中の工程での処理結果が当該処理中の工程に設定された条件を満たした以降に次の段階の工程が動作し、段階を経るに従って、消費電流が大きく、かつ、前記入力された音が音声であるかの判定確度の上がる処理に移行し、それぞれの工程における処理において、その工程に設定された条件が満たされない場合は、それぞれの工程を非動作状態に戻し、前記複数段階の工程における処理により、前記入力された音が音声であると判定された場合は、前記音入力手段の動作状態を維持し、前記入力された音の認識処理へ移行することを特徴とする。

【００１２】

また、本発明の認識対象音声検出方法は、音入力手段に入力された音声を認識してその認識結果に対して何らかの動作を行う音声認識装置における認識対象音声検出方法であって、前記音入力手段を間欠駆動し、間欠駆動される音入力手段が動作状態の間に入力された音のレベルを検出し、そのレベルの大きさから音の有無を判定し、音が無いと判定した場合は、非動作状態に戻る第１の処理工程と、前記第１の処理工程で音が有りと判定された以降に動作を開始し、前記入力された音が雑音であるか音声らしき音であるかを大まかに判定し、音声らしき音ではないと判定した場合は、非動作状態に戻る第２の処理工程と、前記第２段階の処理工程で音声らしき音と判定された以降に動作を開始し、前記音声らしき音と判定された音が音声であるか否かを判定し、音声であると判断した場合は、前記音入力手段の動作状態を維持し、前記音声であると判断された音の音声特徴データを認識部側に渡し、音声でないと判断した場合には、非動作状態に戻る第３の処理工程と、を有したことを特徴とする。

【００１３】

そして、前記第１の処理工程は、前記音入力手段が動作状態の間に入力された音の平均パワーを求め、当該求められた平均パワーと基準レベルとを比較して音の有無を判定し、音が無いと判定した場合は、非動作状態に戻るようにしてもよく、また、前記音入力手段が動作状態の間に入力された音を、人間の音声の周波数帯域を含む周波数帯域とそれ以外の周波数帯域に分けて少なくとも一方の周波数帯域の平均パワーを求め、当該求められた平均パワーの値を基に音を判定し、人間の音声ではないと判定した場合は、非動作状態に戻るようにしてもよい。さらに、これらを組み合わせるようにしてもよい。

【００１４】

また、前記第２の処理工程は、前記第１の処理工程での設定された条件を満たした音信号に対し、当該音信号の継続時間を計測し、当該計測した継続時間を基に音声らしき音か否かを判定し、音声らしき音ではないと判定した場合は、非動作状態に戻るようにしてもよく、また、前記第１の処理工程での設定された条件を満たした音信号に対し、当該音信号の所定時間内における零交差数を計測し、当該計測した零交差数を基に音声らしき音か否かを判定し、音声らしき音ではないと判定した場合は、非動作状態に戻るようにしてもよい。さらに、これらを組み合わせるようにしてもよい。　また、前記第３の処理工程は、前記第２の処理工程での設定条件を満たした音信号に対し、音声特徴抽出処理を行い、前記音声特徴抽出処理により抽出された音声特徴データを基に、入力音が音声であるか否かを判断して音声であると判断した場合に、前記音声であると判断した音の特徴データを認識部側に渡し、音声でないと判断した場合は、非動作状態に戻るようにしてもよい。

【００１５】

さらに、前記認識部では、設定されたキーワードを含む音声特徴データのみを認識処理するようにしてもよい。

【００１６】
また、本発明の認識対象音声検出装置は、音入力手段に入力された音声を認識してその認識結果に対して何らかの動作を行う音声認識装置における認識対象音声検出装置であって、前記音入力手段を間欠駆動する間欠駆動制御手段と、前記間欠駆動制御手段により間欠駆動される音入力手段が動作状態の間に入力された音に対し、前記入力された音が音声であるかを判定する処理を、複数段階に分けて段階的に行うそれぞれの処理手段を有し、現在処理中の処理手段による処理結果が、当該処理中の処理手段に設定された条件を満たした以降に、次の段階の処理手段が動作し、段階を経るに従って、消費電流が大きく、かつ、前記入力された音が音声であるかの判定確度の上がる処理に移行し、それぞれの処理手段における処理において、その処理手段に設定された条件が満たされない場合は、それぞれの処理手段を非動作状態に戻し、前記複数段階の処理手段における処理により、前記入力された音が音声であると判定された場合は、前記音入力手段の動作状態を維持し、前記入力された音の認識処理へ移行するようになっていることを特徴とする。

【００１７】
また、本発明の認識対象音声検出装置は、音入力手段に入力された音声を認識してその認識結果に対して何らかの動作を行う音声認識装置における認識対象音声検出装置であって、前記音入力手段を間欠駆動する間欠駆動制御手段と、前記間欠駆動制御手段により間欠駆動される音入力手段が動作状態の間に入力された音のレベルを検出し、前記検出した音のレベルの大きさから音の有無を判定し、音が無いと判定した場合は、非動作状態に戻る入力レベル判定手段と、前記入力レベル判定手段で音が有りと判定された以降に動作を開始し、前記入力された音が雑音であるか音声らしき音であるかを大まかに判定し、音声らしき音でないと判定した場合は、非動作状態に戻る音判定手段と、前記音判定手段で前記入力された音が音声らしき音と判定された以降に動作を開始し、前記音声らしき音と判定された音が音声であるか否かを判定し、音声であると判定した場合は、前記音声であると判定された音の音声特徴データを認識部側に渡し、音声でないと判断した場合には、非動作状態に戻る音声判定手段と、を有し、前記音判定手段が音声であると判定した場合には、前記音入力手段の動作状態を維持するようになっていることを特徴とする。

【００１８】
そして、前記入力レベル判定手段は、前記音入力手段が動作状態の間に入力された音の平均パワーを求め、当該求められた平均パワーと基準レベルとを比較して音の有無を判定し、音が無いと判定した場合は、非動作状態に戻るようにしてもよく、前記音入力手段が動作状態の間に入力された音を、人間の音声の周波数帯域を含む周波数帯域とそれ以外の周波数帯域に分けて少なくとも一方の周波数帯域の平均パワーを求め、当該求められた平均パワーの値を基に音を判定し、人間の音声ではないと判定した場合は、非動作状態に戻るようにしてもよい。さらに、これらを組み合わせるようにしてもよい。

【００１９】
また、前記音判定手段は、前記入力レベル判定手段での設定された条件を満たした音信号に対し、前記音信号の継続時間を計測し、前記計測した継続時間を基に音声らしき音か否かを判定し、音声らしき音ではないと判定した場合は、非動作状態に戻るようにしてもよく、前記入力レベル判定手段での設定された条件を満たした音信号に対し、前記音信号の所定時間内における零交差数を計測し、前記計測した零交差数を基に音声らしき音か否かを判定し、音声らしき音ではないと判定した場合は、非動作状態に戻るようにしてもよい。さらに、これらを組み合わせるようにしてもよい。

【００２０】
また、前記音声判定手段は、前記音判定手段での設定された条件を満たした音信号に対し、音声特徴抽出処理を行い、前記音声特徴抽出処理により抽出された音声特徴データを基に入力音が音声であるか否かを判断して音声であると判断した場合に、前記音声であると判断した音の音声特徴データを認識部側に渡し、音声でないと判断した場合は、非動作状態に戻るようにしてもよい。

【００２１】

さらに、前記認識部では、設定されたキーワードを含む音声特徴データのみを認識対象音声として認識処理するようにしてもよい。

【００２２】

本発明は、認識対象の音声を常に待ち受ける状態にしておき、認識対象音声が入力されると、認識結果に対応した動作を行う音声認識装置に適用されることで効果を発揮するものである。この種の装置にあっては、音声を常に待ち状態としておくために消費電流が大きく、乾電池を電源として用いるものにあっては、消費電流を如何に小さく抑えるかということが大きな課題であった。

【００２３】

これを解決するために本発明は、まず、音入力手段を間欠駆動する。具体例としては、たとえば、0.1 秒間を動作状態として音声入力を可能とし、その後の0.4 秒間を非動作状態とするというように動作状態と非動作状態を繰り返す間欠的な音声入力動作を行う。このような間欠駆動を行うことにより、待ち状態における消費電流を小さく抑えることができる。

【００２４】

ただし、間欠駆動することによる問題点もある。たとえば、消費電流を小さく抑えるために、たとえば、駆動時間をごく短い時間（たとえば0.1 秒程度）としたとき、マイクロホンの特性上、正常な音声入力動作を行うことができない。これに対処するために、まず、第１段階の処理として、処理時間が短く、しかも電流消費が小さくて済む音の有無検出だけを行い、この第１段階の処理を通過した音信号に対し、第２段階の処理として、その音がどのような音であるかの判定を行い、音声らしいと判定された場合に、第３段階の処理として、人間の音声であるか否かの判定処理を行うというように、幾つかの工程に分けて段階的な処理を行うようにしている。しかも、工程を経るにしたがって、処理時間と消費電流を要する処理とし、それぞれの工程での条件を満たされない場合は、音声入力手段を非動作状態に戻すようにすることで、無駄な電流消費を抑えることができる。また、第１の工程は平均パワーを算出する手段、基準レベルを記憶する手段、比較手段、さらに、場合によっては、周波数フィルタを加える程度で構成でき、簡単に実現可能である。また、第２の工程も、計時手段、継続時間を測定する手段、さらには、零交差数を計測する手段程度で構成でき、第１の工程同様簡単に実現できる。

【００２５】

なお、第１の工程において、音声の有無を判定する場合、周波数フィルタを用いた処理を行うことにより、人間の音声とは異なる音を、早い処理段階で除去することも可能となる。つまり、まず、人間の音声の周波数範囲内に一定のパワーを有する音が有るか否かを判定することで、人間の音声とは異なった周波数帯域に大きな平均パワーを有する音を処理対象外とすることができ、処理の効率化が図れる。また、入力レベルの大きさから音の有無を判定する処理と、この周波数フィルタを用いた処理とを組み合わせることにより、より一層、処理の効率化が図れる。

【００２６】

また、第２の工程において、零交差数を計測する処理を行うことで、効率よく人間の音声らしき音かそれ以外の音かを判定することができる。また、この零交差数を計測する処理と、所定レベル以上の信号の継続時間を判定する処理を組み合わせることにより、より一層、高精度で効率的な音判定が可能となる。

【００２７】

また、第３の工程は、認識装置がもともと持っている音声特徴抽出手段を用いることで実現でき、この処理により人間の音声であるか否かの判定を高精度に行うことができる。

【００２８】

さらに、予め設定されたキーワードを含む音声特徴データのみを認識対象音声として受け付けるようにすることにより、無駄な認識動作を行わなくて済み、これによっても消費電流を小さく抑えることができる。

【００２９】

【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。
【００３０】
図１は本発明の実施の形態を説明するブロック図であり、たとえばコンデンサマイクロホンなどの音入力手段１、音入力手段１から入力された音のレベルが一定以上あるか否かを判定する入力レベル判定手段２、入力レベル判定手段２で一定以上のレベルがあると判定された音が音声らしきものであるかそれ以外の雑音であるか否かを判定する音判定手段３、音判定手段３で音声らしきものであると判定された場合、それが音声であるか否かを判定する音声判定手段４、音声判定手段４で音声であると判定された場合、その音声に対して認識動作を行う音声認識手段５、間欠駆動制御手段６などから構成されている。なお、この間欠駆動制御手段６は、前記音入力手段１に対しては、間欠駆動信号（これについては後述する）を与えるとともに、他の手段に対してはそれぞれの手段が動作を行うときに動作電圧を与える。
【００３１】
このような構成において、その動作を説明する。間欠駆動制御手段６は、音入力手段１に対して間欠駆動信号を与え、これにより、音入力手段１は周期的に動作状態と非動作状態となる。このように、音声入力手段１が間欠駆動する場合、幾つかの問題点がある。
【００３２】
すなわち、音入力手段１が動作状態となる頻度が少ないと、認識すべき音声を入力し損なう可能性がある。逆に、頻度が高すぎると電流消費が大きくなり、低消費電流化に支障がでることになる。これらの点に対しては、ある程度の頻度を有し、かつ、動作状態の時間を短くすることで対処できる。
【００３３】
しかし、動作状態の時間を短くしすぎると、音入力手段１の特性上の問題点がある。たとえば、音入力手段１としてコンデンサマイクロホンを用いた場合、入力した音信号を安定した音信号（たとえば認識処理を行うための特徴抽出が可能な音信号）として取り出すには、通常、秒単位の時間を要する。
【００３４】
これらの点を考慮して、本発明では、音入力手段１を間欠駆動させ、かつ、音入力手段１が動作状態のときに取り込んだ音信号に対して複数段階の工程に分けて、順次、消費電流が大きく、認識対象音声か否かの判定確度の上がる処理時間の長い処理を行うようにする。以下、この具体的な処理について説明する。
【００３５】
この実施の形態においては、前述した点を考慮して、音入力手段１に対し、たとえば、0.1 秒間動作可能状態とし、その後の0.4 秒は動作を休み、その後、再び、0.1 秒間動作可能状態とし、その後の0.4 秒は動作を休むというような間欠駆動信号を間欠駆動制御手段６から出力する。
【００３６】
したがって、音声入力手段１は、間欠駆動制御手段６によって間欠的に設定されるる0.1 秒間の動作可能状態のときにのみ音の入力を可能とし、それ以外では音声入力動作やその他の動作を行わない状態（これをここではスリープ状態という）となる。
【００３７】
そして、たとえば、ある時刻において音信号が存在し、このとき、音入力手段１が動作可能状態となっていれば、その音は音入力手段１により取り込まれる。この音入力手段１に入力された音信号は、入力レベル判定手段２で、入力レベルの判定が行われる。つまり、この段階では音の有無だけの判定を行う。
【００３８】
この入力レベル判定手段２による音の有無検出は、様々な手法により行うことができる。たとえば、図２に示すような例がある。この図２で示す例は、平均パワー算出部２１１、基準レベル記憶部２１２、比較部２１３、入力音判定結果出力部２１４で構成され、音入

力手段１で入力された音信号から平均パワーを算出して、その平均パワーを基準レベルと比較し、その比較結果に基づいて入力音判定結果を出力する。

【００３９】

なお、音入力手段１としてコンデンサマイクロホンを用いた場合、前述したように、入力した音信号を安定した音信号とするには、通常、秒単位の時間を要するが、このように、単に音があるか無いかを判定するだけの処理を行うには、0.1 秒程度の時間で実用的には十分である。

【００４０】

以上の処理は、本発明の第１段階の処理であり、図３のフローチャートのステップｓ１～ｓ３の処理である。つまり、スリープ状態（ステップｓ１）において、間欠駆動制御手段６から動作開始信号が入ると、音入力手段１が動作状態となり、所定レベル以上の音信号があるか否かを判定する（ステップｓ２，ｓ３）。そして、所定レベル以上の音信号が存在すると判定されたときは、次の第２段階の処理に移り、もし、所定レベル以上の音信号が無ければ、音は無しと判断してスリープ状態に戻る。

【００４１】

音が有りと判断された場合は、第２段階の処理として、音判定手段３によりその音が音声らしき音であるか雑音であるかを判定する。この音声らしき音であるか雑音であるかを判定する手段としては幾つか考えられるが、ここでは、その一例として図４に示すように、所定以上のレベルの音の継続時間を調べて突発的な雑音であるか否かを判定する。

【００４２】

図４に示す音判定手段３は継続時間判定部３１、計時部３２、継続時間記憶部３３、音判定結果出力部３４などから構成されている。このような構成において、入力レベル判定手段２にて所定レベル以上と判定された信号がどの程度継続しているかを計時部３２からの時間信号を用いて計時し、継続時間記憶部３３に記憶されている時間に基づいてその入力音が音声らしき音かそれ以外の突発的な雑音かを判定する。

【００４３】

つまり、所定レベル以上の入力音の継続時間が継続時間記憶部３３に記憶されている時間より短い場合は、少なくとも音声ではなく、たとえば、ドアを閉めたときの音などの突発的な雑音であると判定する。

【００４４】

以上の第２段階の処理は、図３のフローチャートのステップｓ４，ｓ５の処理である。つまり、第１段階の処理（音が有るか否かの判定処理）において、音が有りと判定された場合、まず、音判定処理として、前述したような所定レベル以上の音の継続時間を調べ（ステップｓ４）、その時間にもとづいて入力音は雑音であるか否かを判定する（ステップｓ５）。ここで、突発的な雑音でない、つまり、音声の可能性があると判定された場合は、次の第３段階に処理に移り、もし、突発的な雑音であると判定された場合はスリープ状態に戻る。

【００４５】

この第２段階の処理において、入力音声が突発的な雑音ではなく、音声である可能性があると判定された場合、第３段階の処理として、音声判定手段４によりその音声らしき音が人間の音声であるか否かを判定する。この音声判定手段４による音声判定処理について以下に説明する。

【００４６】

この音声判定手段４が行う処理は、まず、音声らしき音が人間の音声かそれ以外の音かを区別することが必要であるが、これに対しては、入力音を特徴抽出処理（たとえばＬＰＣ分析）し、その分析結果に基づいて人間の音声であるか否かを判定する。具体的には、人間の音声生成機構のモデル化による特徴抽出処理を行って、その誤差を求め、誤差の大きさから人間の音声か否かを判定する。たとえば、第２段階の処理で突発的な雑音ではなく音声らしき音と判定された場合でも、音信号をＬＰＣ分析による誤差を求めることで明確に判断できる。なお、この音声判定手段４は、音声認識装置がもともと持っている特徴分

析手段により行うことができることは勿論である。

【００４７】

以上の第３段階の処理は、図３のフローチャートのステップｓ６，ｓ７の処理である。つまり、第２段階の処理（音声らしい音か否かの判定処理）において、音声らしいと判定された場合、音声判定処理として、特徴抽出を行い（ステップｓ６）、その結果に基づいて人間の音声であるか否かを判定し、人間の音声であると判定した場合は認識対象音声として（ステップｓ７）、音声認識手段５にその特徴データを送り、認識処理に移る。もし、ステップｓ７で、人間の音声でないと判定された場合は、認識対象音声でないとしてスリープ状態に戻る。また、音声認識手段５により認識処理を行い（ステップｓ８）、認識処理が終了するとスリープ状態に戻る。

【００４８】

以上のように本発明は、第１～第３の三段階の処理を経て、入力された音が人間の音声であると判定された場合に初めて認識処理に入るようにしている。

【００４９】

つまり、第１段階では、間欠的に入力音を検出してレベルの大きさから入力音が有るか否かだけの処理を行い、所定レベル以上の入力音が存在した場合に、第２段階の処理を行う。そして、第２段階では、所定レベル以上の入力音が突発的な雑音であるのか音声らしき音であるのかを判定し、音声らしき音である場合にのみ第３段階の処理に入る。この第３段階の処理では、音声らしき音が人間の音声であるか否かを判定し、人間の音声である場合にそれを認識対象として音声認識手段５に特徴データを渡すようにする。

【００５０】

なお、第１～第３の三段階の処理に要する時間は、たとえば「今、何時」というような認識対象音声の発話時間に比べると、ごく短い時間であるので、認識対象音声に対して認識処理を行う上で、実用的には殆ど支障はない。

【００５１】

以上説明したように、本発明では、第１段階の動作における音の待ち状態は、この実施の形態では、0.５秒間のうち、0.1 秒の動作を行う間欠動作であるので、常に入力音声を待つ状態としておく場合に比べ、１／５の消費電流ですむ。

【００５２】

ちなみに、単１の乾電池は単３の乾電池の4.5倍ほどの容量があるので、消費電流が１／５となれば、単１と同じ電池寿命を得ようとした場合、単３電池で可能となる。

【００５３】

また、第１段階から順に段階を経るごとに、消費電流が大きく、動作時間の長い処理に移るようにし、１つの段階に設定された条件を満たしたとき、次の段階に移り、ある段階で条件が満たされないときはスリープ状態に戻るので、無駄な処理を行わなくて済み、これによっても消費電流を抑えることができる。特に、第３段階以降の処理は、実質的な音声認識処理であり、装置としてフル稼働に近い状態となるが、この第３段階に達するまでに、条件が満たされないときは、スリープ状態に戻るため、無駄な認識動作を行わなくて済む。

【００５４】

なお、以上説明した例は１つの実施の形態であり、この実施の形態に限定されるものではない。たとえば、間欠駆動制御手段６による音入力手段１に対する駆動は、上述の実施の形態では、0.1 秒間を動作状態とし0.４秒休むというように、0.5 秒間のうち0.1 秒の動作を行うようにしているが、この動作頻度は任意に設定できるものであり、その装置の特性などを考慮して最も適当と思われる動作間隔を設定できるものである。ただし、頻度が少なすぎると、入力音の検出ミスが発生しやすいので、ある程度の頻度は必要である。

【００５５】

また、第１段階の処理における入力レベル判定手段２は、図２で示したものでなく、たとえば、図５（ａ）あるいは図５（ｂ）のような構成としてもよい。

【００５６】

図５（ａ）は、低域通過フィルタ２１５、平均パワー算出部２１６、基準レベル記憶部２１７、比較部２１８、入力レベル判定結果出力部２１９から構成されている。低域通過フィルタ２１５は、ここでは、４ＫＨｚ以下の周波数成分を通過させるもので、４ＫＨｚ以下の周波数成分の平均パワーを平均パワー算出部２１６で算出し、その平均パワーと、基準レベル記憶部２１７に記憶されている基準レベルとを比較部２１８で比較し、その比較結果をもとに入力レベル判定結果を出力する。

【００５７】

ここで、周波数の基準を４ＫＨｚとしたのは、人間の音声は殆どが４ＫＨｚ以下であるためである。これにより、４ＫＨｚ以下の周波数成分の平均パワーが、基準レベルより大きい場合には、人間の声である可能性もあると判定できる。言い換えれば、４ＫＨｚ以上に大きな値の平均パワーを有する音は人間の音声とは異なる音であるとみなすことができるので、このような音は処理対象から除去する。したがって、まず、人間の音声の周波数範囲内に一定のパワーを有する音を判定結果として取り出し、この音に対して第２段階以降の処理を行う。

【００５８】

また、図６（ｂ）は、４ＫＨｚ以下の周波数成分を通過させる低域通過フィルタ２１５と、４ＫＨｚより高い周波数成分を通過させる高域通過フィルタ２２０と、４ＫＨｚ以下の周波数成分の平均パワーを算出する平均パワー算出部２２１、４ＫＨｚより高い周波数成分の平均パワーを算出する平均パワー算出部２２２、これらの平均パワーの差あるいは比をとる比較部２２３、入力レベル判定結果出力部２２４から構成されている。

【００５９】

このように、高域と低域の２種類の周波数帯域フィルタを設け、その差あるいは比を求めることにより、人間の音声とその他の雑音とをさらに精度よく区別することができる。たとえば、４ＫＨｚより高い周波数成分の平均パワーが、４ＫＨｚ以下の周波数成分の平均パワーよりきわめて大きければ、それは人間の音声ではなく雑音の可能性が高いと判定でき、また、逆に、４ＫＨｚ以下の周波数成分の平均パワーが４ＫＨｚより高い周波数成分の平均パワーよりきわめて大きければ、人間の音声である可能性が有ると判定できる。また、低周波成分と高周波成分の両方にまんべんなくパワーが存在する場合は、両者の比は小さくなり、この場合も、人間の音声以外の雑音であると判定できる。

【００６０】

このように、２種類の周波数帯域のフィルタを設け、それぞれのフィルタを通過する周波数成分の平均パワーの大きさに基づいた判定を行うことにより、この第１段階の処理においても、人間の音声かそれ以外の音声かを大まかに判定することができ、その後の処理を、より一層、効率的に行うことが可能となる。

【００６１】

なお、入力レベル判定手段２としては、図２で示した手段と図５（ａ），（ｂ）のいずれかの手段とを組み合わせるようにしてもよい。たとえば、図２で示したものと図５（ａ）のものとを組み合わせた場合、まず、入力音声の平均パワーを算出して基準レベルと比較し、基準レベル以上の場合に、低域通過フィルタを通して、４ＫＨｚ以下の周波数成分の平均パワーを算出し、その平均パワーと、基準レベルとを比較し、その比較結果をもとに入力レベル判定結果を出力するようにしてもよい。

【００６２】

また、音判定手段３は、前述の実施の形態では、入力レベル判定手段２からの入力レベル判定結果を受けて、所定以上のレベルの音の継続時間を調べて、突発的な雑音を処理対象外とすることで、音声らしき音を検出するようにしたが、この音判定手段３は、たとえば、図６に示すような零交差数計測を行うことで、音声らしき音を検出するようにしてもよい。図６は零交差数計測部３５、計時部３６、音判定結果出力部３７で構成され、音入力手段１に入力される音信号の零交差数を調べることにより、その音が人間の音声であるかそれ以外の音で有るかを大まかに判定する。

【００６３】

つまり、ある一定時間における人間の音声の零交差数は予めわかっているので、入力音に対してある一定時間内の零交差数をカウントし、その零交差数を基に音判定を行う。これにより、第１段階における設定条件を満たした音であっても、たとえば、電話の呼び出し音、チャイムの音、楽器の音、機械音などの音を人間の音声と区別することができ、より人間の音声らしき音のみを判定することができる。

【００６４】

また、音判定手段３としては、前述の継続時間を調べて突発的な雑音を除去する手段と、図６で示した手段とを併用するようにしてもよい。たとえば、まず最初に、継続時間を判定し、所定の時間以上継続する音であると判定した場合に、入力音声の零交差数を調べ、その零交差数により音声らしき音か雑音かの判定を行うようにする。これにより、入力音が人間の音声らしき音であるか否かを高い精度で判定できる。

【００６５】

また、第３段階の処理として、音声判定手段４が行う処理は、ＬＰＣ分析などの音声特徴抽出により非音声を除去する処理であるが、たとえば、テレビジョンやラジオから流れてくる人間の音声は認識対象の音声と判断してしまうことになる。このような認識対象ではない人間の音声特徴データが音声認識部５に与えられると、音声認識部５ではその音声に反応して訳の分からない応答をすることがある。このような認識対象音声以外の音声を排除するために、その後の音声認識手段５における認識処理をキーワードを用いて認識を行うようにしてもよい。

【００６６】

つまり、音声認識手段５における認識可能な登録単語の１つとしてキーワードを予め登録しておき、そのキーワードを含んだ音声を入力することで認識動作が可能とするような設定としておく。

【００６７】

たとえば、時刻を問い合わせると現在時刻を応答する時計を考えた場合、キーワードとしてたとえば「太郎」を予め登録しておき、時刻を問い合わせるときに、単に「今何時」というのではなく、たとえば、「太郎、今何時」というように、キーワードを含んだ内容の問いかけを行うようにする。装置側では、キーワードが含まれている場合だけ認識対象音声として受け付けるようにする。これにより、キーワードを含まない音声は、認識対象音声として受け付けないので、前述したように、テレビジョンやラジオから流れてくる人間の音声に装置が反応して訳の分からない応答をするというようなことがなくなり、これによっても、無駄な電流消費を抑えることができる。

【００６８】

なお、以上説明した実施の形態は、本発明の好適な実施の形態の例であるが、これに限定されるものではなく、本発明の要旨を逸脱しない範囲で、種々変形実施可能である。

【００６９】

なお、本発明の処理を行う処理プログラムは、フロッピィディスク、光ディスク、ハードディスクなどの記憶媒体に記憶させておくことができ、本発明は、それらの記憶媒体をも含むものであり、また、ネットワークからデータを得る形式でもよい。

【００７０】

以上説明したように、本発明によれば、音入力手段を間欠駆動させ、音入力手段が動作状態のときのみに音声入力動作を行うことにより、待ち状態における消費電流を小さく抑えることができる。

【００７１】

また、本発明では、音入力手段が動作状態のときの処理を幾つかの段階に分けて行う。まず、処理時間が短く、しかも、電流消費が小さくて済む音の有無検出を第１段階の処理として行い、この第１段階の処理を通過した音信号に対し、その音がどのような音であるかの判定を第２段階の処理として行い、この第２段階の処理により音声らしいと判定された場合に、第３段階の処理として、人間の音声であるか否かの判定処理を行うというように、幾つかの工程に分けて処理を行うようにしている。しかも、工程を経るにしたがって、

処理時間と消費電流を要する処理とし、それぞれの工程での条件が満たされない場合は、装置を非動作状態に戻し、音声入力手段のみが間欠駆動するモードに戻すようにしている。

【００７２】

このように、工程を経るにしたがって、処理時間と消費電流を要する処理とすることにより、音入力手段を間欠駆動することによる様々な問題点に対応することができ、しかも、消費電流を大幅に抑えることが可能となる。

【００７３】

これにより、電源として電池を使用する機器の場合、たとえば、単３電池で単１電池と同じ寿命を得ることも可能となり、同じ寿命を得るのに、電池容量を小さなものとすることができ、装置の小型化と軽量化が図れる。また、電池を装置に付加して販売する場合は、電池容量が小さい分、装置の販売価格の低廉化にも寄与することができなど、種々の効果が得られる。

【図面の簡単な説明】

【図１】本発明の実施の形態の基本的な構成を示すブロック図。

【図２】図１で示した入力レベル判定手段の一例を示す図。

【図３】本発明の実施の形態の処理を説明するフローチャート。

【図４】図１で示した音判定手段の一例を示す図。

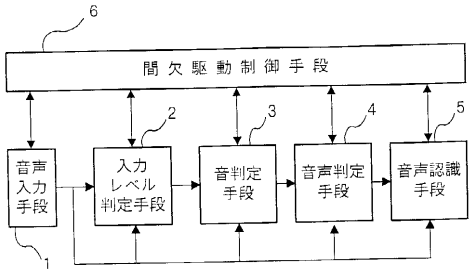
【図５】図１で示した入力レベル判定手段の他の例を示す図。

【図６】図１で示した音判定手段の他の例を示す図。

【符号の説明】

１　音入力手段
２　入力レベル判定手段
３　音判定手段
４　音声判定手段
５　音声認識手段
６　間欠駆動制御手段
２１１，２１６，２２１，２２２　平均パワー算出部
２１２，２１７　基準レベル記憶部
２１３，２１８，２２３　比較部
２１４，２１９，２２４　入力レベル判定結果出力部
２１５　低域通過フィルタ
２２０　高域通過フィルタ
３１　継続時間判定部
３２，３６　計時部
３３　継続時間記憶部
３４，３７　音判定結果出力部
３５　零交差数計測部

【図１】



【図２】

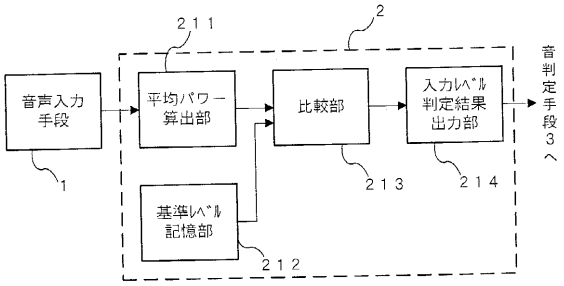
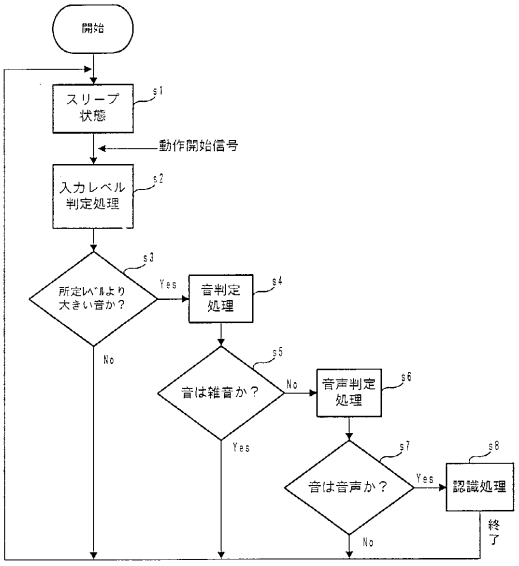
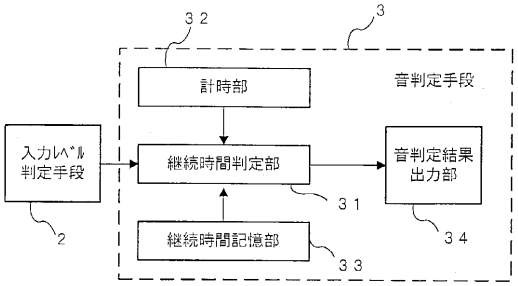


図２

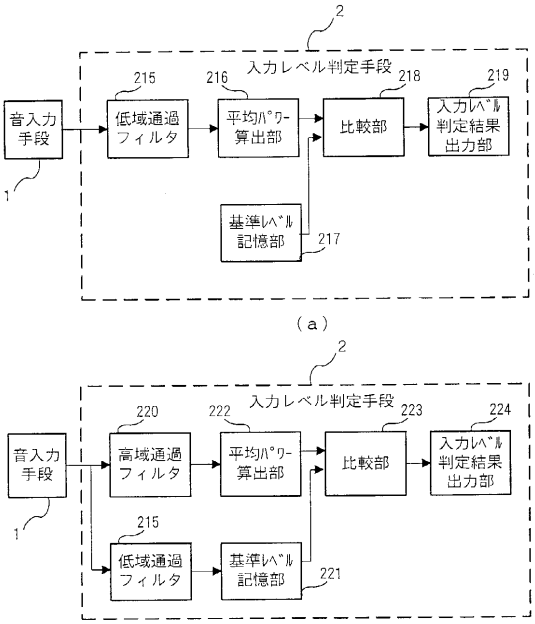
【図３】



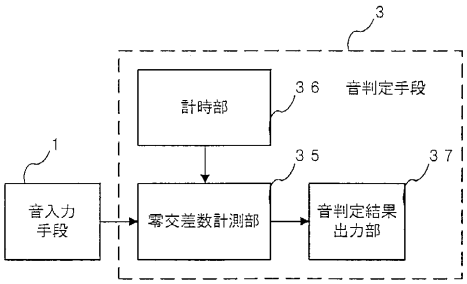
【図４】



【図５】



【図６】



フロントページの続き

審査官　荏原　雄一

(56)参考文献　特開平０９－０６２２９３（ＪＰ，Ａ）
特開昭６２－２６９４２７（ＪＰ，Ａ）
特公昭５３－０３４８０５（ＪＰ，Ｂ２）

(58)調査した分野(Int.Cl.[7]，ＤＢ名)
G10L 11/02
G10L 15/04-15/28
JSTPlusファイル(JOIS)
IEEE Xplore