(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第5053950号
(P5053950)
(45)発行日　平成24年10月24日(2012.10.24)　(24)登録日　平成24年8月3日(2012.8.3)

(51)Int.Cl.　F I
H04N 5/225 (2006.01)　H04N 5/225 F
H04N 101/00 (2006.01)　H04N 101:00

請求項の数 14　(全 37 頁)

(21)出願番号　特願2008-194800 (P2008-194800)
(22)出願日　平成20年7月29日 (2008.7.29)
(65)公開番号　特開2010-34841 (P2010-34841A)
(43)公開日　平成22年2月12日 (2010.2.12)
審査請求日　平成23年7月12日 (2011.7.12)

(73)特許権者　000001007
キヤノン株式会社
東京都大田区下丸子３丁目３０番２号
(74)代理人　100126240
弁理士　阿部　琢磨
(74)代理人　100124442
弁理士　黒岩　創吾
(72)発明者　山本　寛樹
東京都大田区下丸子３丁目３０番２号キヤノン株式会社内

審査官　深沢　正志

最終頁に続く

(54)【発明の名称】情報処理方法、情報処理装置、プログラムおよび記憶媒体

(57)【特許請求の範囲】
【請求項１】
情報処理装置が行う情報処理方法であって、
第１の検出手段が、予め設定された基準を満たす音の開始を検出する第１の検出工程と、
第１の取得手段が、前記開始の検出に応答して第１の画像データを取得する第１の取得工程と、
第１の記憶手段が、前記第１の画像データをメモリに記憶する第１の記憶工程と、
第２の検出手段が、前記音の終了を検出する第２の検出工程と、
第２の取得手段が、前記終了の検出に応答して第２の画像データを取得する第２の取得工程と、
第２の記憶手段が、前記第２の画像データを前記メモリに記憶する第２の記憶工程と、
決定手段が、前記音に含まれる意味に応じて、前記第１の画像データまたは前記第２の画像データのいずれかを保存する対象のデータとして決定する決定工程とを有することを特徴とする情報処理方法。
【請求項２】
情報処理装置が行う情報処理方法であって、
第１の検出手段が、予め設定された基準を満たす音の開始を検出する第１の検出工程と、
第２の検出手段が、前記音の終了を検出する第２の検出工程と、

取得手段が、前記開始または前記終了の検出に応答して画像データを取得する取得工程と、
記憶手段が、前記取得工程で取得した前記画像データをメモリに記憶する記憶工程と、
決定手段が、前記音に含まれる意味に応じて、前記メモリに記憶した前記画像データを保存する対象のデータとして決定する決定工程とを有することを特徴とする情報処理方法。

【請求項３】
更に、消去手段が、前記決定工程で、保存する対象のデータとして決定されなかった画像データを前記メモリから消去する消去工程を有することを特徴とする請求項１または請求項２に記載の情報処理方法。

【請求項４】
更に、保存手段が、前記保存する対象のデータとして決定された前記画像データを第２のメモリに保存する保存工程を有することを特徴とする請求項１乃至請求項３のいずれか１項に記載の情報処理方法。

【請求項５】
前記画像を取得する工程は、前記開始を検出した時点または前記終了を検出した時点に実行されることを特徴とする請求項１乃至請求項４のいずれか１項に記載の情報処理方法。

【請求項６】
前記開始を検出した時点で画像データを取得し、前記開始を検出した時点から前記音が予め設定された時間継続しなかった場合、更に、前記メモリから取得した前記画像データを消去する第２の消去工程と、
前記第２の消去工程に続いて、前記第１の検出工程に相当する、第１の再検出工程とを実行することを特徴とする請求項１乃至請求項５のいずれか１項に記載の情報処理方法。

【請求項７】
前記終了を検出した時点で画像データを取得し、前記終了を検出した時点から予め設定された時間に再び予め設定した基準を満たす音を検出した場合、更に、前記メモリか取得したら前記画像データを消去する第３の消去工程と、
前記第３の消去工程に続いて、前記第２の検出工程に相当する、第２の再検出工程とを実行することを特徴とする請求項１乃至請求項６のいずれか１項に記載の情報処理方法。

【請求項８】
前記画像を取得する工程は、前記開始を検出した時点から予め設定した遅延時間が経過した時点または前記終了を検出した時点から予め設定した遅延時間が経過した時点に実行されることを特徴とする請求項１乃至請求項７のいずれか１項に記載の情報処理方法。

【請求項９】
前記予め設定された基準とは、一定以上の音量を有するか否かであることを特徴とする請求項１乃至請求項８のいずれか１項に記載の情報処理方法。

【請求項１０】
前記音を音声認識することによって前記意味を特定することを特徴とする請求項１乃至請求項９のいずれか１項に記載の情報処理方法。

【請求項１１】
予め設定された基準を満たす音の開始を検出する第１の検出手段と、
前記開始の検出に応答して第１の画像データを取得する第１の取得手段と、
前記第１の画像データをメモリに記憶する第１の記憶手段と、
前記音の終了を検出する第２の検出手段と、
前記終了の検出に応答して第２の画像データを取得する第２の取得手段と、
前記第２の画像データを前記メモリに記憶する第２の記憶手段と、
前記音に含まれる意味に応じて、前記第１の画像データまたは前記第２の画像データのいずれかを保存する対象のデータとして決定する決定手段とを有することを特徴とする情報処理装置。

【請求項１２】
予め設定された基準を満たす音の開始を検出する第１の検出手段と、
前記音の終了を検出する第２の検出手段と、
前記開始または前記終了の検出に応答して画像データを取得する取得手段と、
前記取得手段で取得した前記画像データをメモリに記憶する記憶手段と、
前記音に含まれる意味に応じて、前記メモリに記憶した前記画像データを保存する対象のデータとして決定する決定手段とを有することを特徴とする情報処理装置。
【請求項１３】
コンピュータを、請求項１１又は請求項１２に記載の情報処理装置が有する各手段として機能させるためのプログラム。
【請求項１４】
請求項１３に記載のプログラムを記憶したコンピュータ読み取り可能な記憶媒体。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、音で撮像を指示する技術に関する。
【背景技術】
【０００２】
従来、一定以上の音量を検知すると撮像動作を実行する機能（以下、音量検知シャッターとする）を備えたカメラが知られている（例えば、特許文献１）。この機能を利用すると、発声のタイミングに合わせて撮像することが可能となる。
【０００３】
また、撮像を指示する声を認識すると撮像動作を実行する機能（以下、音声認識シャッターとする）を備えたカメラが知られている（例えば、特許文献２）。この機能を利用すると、ユーザが撮像を所望して発声した場合に撮像することが可能となる。なお、音声認識シャッターを利用して撮像する場合、ユーザが撮像を指示する声を発しても、音声コマンドの発声が完了するまではカメラの撮像動作は実行されない。よって、所望する撮像のタイミングを逃してしまうことがある。
【特許文献１】特開平１１－１９４３９２号公報
【特許文献２】特開２００６－１８４５８９号公報
【発明の開示】
【発明が解決しようとする課題】
【０００４】
従来の音量検知シャッターを利用して撮像する場合には、音の発声タイミングに連動して撮像動作を実行できる。しかし、この場合には、例えば大きな雑音等、目的の声以外を検知した場合にも撮像動作を実行してしまうため、不要な画像を保存してしまうという課題がある。
【０００５】
例えば、“撮影”という発声に基づいてユーザが所望するタイミングで撮像する工程と、“消去”という音声コマンドに基づいて既に撮像した画像を消去する工程を備えることにより、上記課題を解決できる。しかしながら、２種類の音声コマンドを入力する作業は効率がよくない。
【０００６】
本発明は係る従来例を鑑みてなされたものであり、単一の音声コマンドに基づいて、音が入力されたタイミングを反映した撮像で得られた画像であって、かつユーザが所望する画像を効率良く保存することを主な目的とする。
【課題を解決するための手段】
【０００７】
上記目的を達成するための情報処理方法の１つとして、情報処理装置が行う情報処理方法であって、第１の検出手段が、予め設定された基準を満たす音の開始を検出する第１の

検出工程と、第１の取得手段が、前記開始の検出に応答して第１の画像データを取得する第１の取得工程と、第１の記憶手段が、前記第１の画像データをメモリに記憶する第１の記憶工程と、第２の検出手段が、前記音の終了を検出する第２の検出工程と、第２の取得手段が、前記終了の検出に応答して第２の画像データを取得する第２の取得工程と、第２の記憶手段が、前記第２の画像データを前記メモリに記憶する第２の記憶工程と、決定手段が、前記音に含まれる意味に応じて、前記第１の画像データまたは前記第２の画像データのいずれかを保存する対象のデータとして決定する決定工程とを有することを特徴とする。

【発明の効果】

【０００８】

本発明によれば、単一の音声コマンドに基づいて、音が入力されたタイミングを反映した撮像で得られた画像であって、かつユーザが所望する画像を効率良く得ることができる。

【発明を実施するための最良の形態】

【０００９】

以下、本発明に好適な実施形態について、図面を参照しながら説明していく。

【００１０】

（第１の実施形態）

図１は第１の実施形態に係る情報処理装置の構成の一例であるデジタルカメラを示す機能ブロック図である。

【００１１】

図１においてデジタルカメラ２００は、制御部１０１、操作部１０２、撮像部１０３、メモリ（画像記憶用）１１０、記憶媒体（画像記憶用）１１１を備える。

【００１２】

また、デジタルカメラ２００は、マイク１１２、メモリ（音声認識データ用）１１３、メモリ（認識結果制御テーブル用）１１４、ディスプレイ１１５を備える。

【００１３】

（各部の説明）

制御部１０１は、操作部１０２、撮像部１０３、メモリ（画像記憶用）１１０、記憶媒体（画像記憶用）１１１、マイク１１２、メモリ（音声認識データ用）１１３、メモリ（認識結果制御テーブル用）１１４、ディスプレイ１１５の動作を制御する。

【００１４】

尚、制御部１０１における処理は後述する。

【００１５】

また、制御部１０１は、ＣＰＵ（中央演算装置）、ＲＯＭ（Ｒｅａｄ　Ｏｎｌｙ　Ｍｅｍｏｒｙ）、ＲＡＭ（Ｒａｎｄｏｍ　Ａｃｃｅｓｓ　Ｍｅｍｏｒｙ）等によって構成される。

【００１６】

また、制御部１０１は、ソフトウエアモジュールとして操作制御部１２２、撮像制御部１２３、画像記憶制御部１０４、音声入力部１０５、音声検出部１０６、音声認識部１０７、認識結果処理部１０８、表示制御部１０９を有する。

【００１７】

操作制御部１２２は、ユーザが操作部１０２に対して行った操作を検知するための部分である。

【００１８】

撮像制御部１２３は、撮像部１０３に撮像動作を実行させるための部分である。

【００１９】

画像記憶制御部１０４は、メモリ（画像記憶用）１１０および記憶媒体（画像記憶用）１１１へのデータの書込み、メモリ（画像記憶用）１１０および記憶媒体（画像記憶用）

１１１に記憶されているデータの読み出し、消去等を制御する。
【００２０】
音声入力部１０５は、マイク１１２を介して入力される音をデジタルの音声信号に変換して出力する部分である。
【００２１】
音声検出部１０６は、音声入力部１０５が変換したデジタルの音声信号をフレーム単位で順次処理し、基準を満たす音の開始および終了を検出する。
【００２２】
尚、ユーザが発声した区間（時間帯）は、基準を満たす音の開始を検出してから基準を満たす音の終了を検出するまでの時間帯を発声区間とする。
【００２３】
尚、フレームとは、時間的に変化する音声信号をほぼ定常とみなせる固定時間長（例えば、２５．６ミリ秒とする）毎に区分するために設けた処理単位である。なお、このフレーム数によって時間を表現することも可能である。
【００２４】
音声認識部１０７は、ソフトウエアモジュールとして音響分析部、探索部を有し、ユーザが発声した区間に含まれるコマンド（いわゆる音声コマンド）を認識する。
【００２５】
尚、コマンドとは音声認識部１０７が認識可能な音のまとまりであり、例えば、“Ｓｈｏｏｔ”等である。
【００２６】
音響分析部は、音声信号をフレーム単位で分析して、例えばＭＦＣＣ（Ｍｅｌ　Ｆｒｅｑｕｅｎｃｙ　Ｃｅｐｓｔｒｕｍ　Ｃｏｅｆｆｉｃｃｉｅｎｔ）等の特徴量のデータを出力する。
【００２７】
探索部は、例えば、Ｖｉｔｅｒｂｉアルゴリズム等の周知のアルゴリズムを用いた探索処理を行い、所定個数のコマンドと、各々のコマンドに対応する認識スコアとを認識結果として出力する。
【００２８】
また、探索部は、探索処理を実行する際、メモリ（音声認識データ用）１１３に含まれる音響モデルと言語モデルとを用いる。
【００２９】
尚、音響モデル、言語モデルの詳細は後述する。
【００３０】
尚、認識スコアとは、音響的な類似度を示す周知の音響スコア、言語モデルから求まる周知の言語スコア、またはこれら２つの重みつき和であってもよい。また、認識結果の確からしさを示す周知の信頼度スコアでもよい。
【００３１】
尚、異なるスコアまたは複数のスコアの用いることで、種々の音の応じた最適な探索処理を実行することが可能となる。
【００３２】
認識結果処理部１０８は、音声認識部１０７が出力した認識結果のデータを取得し、メモリ（認識結果制御テーブル用）１１４に記憶された認識結果制御テーブルを参照して、認識結果に含まれるコマンドに対応する制御を決定する。
【００３３】
尚、本実施形態に利用する認識結果制御テーブルの一例は後述する。
【００３４】
表示制御部１０９は、ディスプレイ１１５に表示する表示内容を制御する。
操作部１０２は、ユーザがデジタルカメラ２００を手動で操作するため部分である。
尚、操作部１０２は、ボタン、スイッチ等によって構成される。

【００３５】
撮像部１０３は、レンズによって結像した像の撮像信号を生成し、生成された撮像信号にＡ／Ｄ変換等の画像処理を施す。
尚、撮像部１０３は、レンズ、撮像センサ等によって構成される。
【００３６】
メモリ（画像記憶用）１１０は、撮像部１０３が撮像した画像の画像データを一時的に記憶する。尚、メモリ（画像記憶用）１１０は、ＲＡＭ等である。
【００３７】
記憶媒体（画像記憶用）１１１は、撮像部１０３が撮像した画像の画像データを最終的に蓄積する。尚、記憶媒体（画像記憶用）１１１は、不揮発性メモリである。
【００３８】
メモリ（画像記憶用）１１０は第１のメモリとして機能し、記憶媒体（画像記憶用）は第２のメモリとして機能する。
【００３９】
マイク１１２は、ユーザの音声入力を受け付け、入力された音声データを音声入力部１０４に出力する。
尚、マイク１１２は、周知のモノラルマイク、ステレオマイク等である。
【００４０】
メモリ（音声認識データ用）１１３は、音声認識の実行に必要なデータと、例えばＨＭＭ（Ｈｉｄｄｅｎ　Ｍａｒｋｏｖ　Ｍｏｄｅｌ）等の周知の音響モデルと、Ｎ－ｇｒａｍ、形態素解析等の周知の言語モデルとを記憶する。
【００４１】
尚、Ｎ－ｇｒａｍとは、語の連鎖確率等を用いて言語の統計的な情報によって構成された言語モデルである。
【００４２】
また、音声認識で受理可能な特定の語や語の接続規則を記述した音声認識文法を言語モデルとして利用してもよい。尚、本実施形態に利用する音声認識文法の一例は、後述する。
【００４３】
また、メモリ（音声認識データ用）１１３は、不揮発性メモリ等である。
メモリ（認識結果制御テーブル用）１１４は、認識結果制御テーブルを格納する。また、メモリ（認識結果制御テーブル用）１１４は、不揮発性メモリである。
尚、本実施形態に利用する認識結果制御テーブルの一例は後述する。
【００４４】
尚、不揮発性メモリとは、周知のハードディスク、コンパクトフラッシュ（登録商標）、ＳＤ（Ｓｅｃｕｒｅ　Ｄｉｇｉｔａｌ）カード等でもよい。
【００４５】
また、不揮発性メモリとは、ＣＤ（Ｃｏｍｐａｃｔ　Ｄｉｓｋ）、ＤＶＤ（Ｄｉｇｉｔａｌ　Ｖｅｒｓａｔｉｌｅ　Ｄｉｓｋ）でもよい。
【００４６】
また、不揮発性メモリとは、ＬＡＮ（Ｌｏｃａｌ　Ａｒｅａ　Ｎｅｔｗｏｒｋ）アダプタ、ＵＳＢ（Ｕｎｉｖｅｒｓａｌ　Ｓｅｒｉａｌ　Ｂｕｓ）アダプタ等のインタフェースを介して情報処理装置１００と接続可能な外部の記憶媒体であってもよい。
【００４７】
ディスプレイ１１５は、撮像部１０３で撮像された画像、メモリ（画像記憶用）１１０、記憶媒体（画像記憶用）１１１等に記憶された画像等を表示する。
【００４８】
また、ディスプレイ１１５は、例えばＬＣＤ（Ｌｉｑｕｉｄ　Ｃｒｙｓｔａｌ　Ｄｉｓｐｌａｙ）や有機ＥＬ（Ｅｌｅｃｔｒｏ－Ｌｕｍｉｎｅｓｃｅｎｃｅ）等である。
【００４９】

（カメラ本体の外観の説明）
図２は、本実施形態で想定されるデジタルカメラの外観を示す図である。尚、図２（Ａ）はデジタルカメラ２００の前面の外観、図２（Ｂ）はデジタルカメラ２００の背面の外観である。
【００５０】
尚、図１と共通の要素には同一の符号を付し、その説明を省略する。
【００５１】
図２において、デジタルカメラ２００は、シャッターボタン２０１、音声シャッター切替えスイッチ２０２、モードダイアル２０３、四方向選択ボタン２０４、決定ボタン２０５、電源ボタン２０６、録音ボタン２０７を備える。これらは、図１の操作部１０２に相当する。
【００５２】
（デジタルカメラ２００の各部の説明）
２０１は、撮像を指示するための操作に用いるシャッターボタンである。
２０２は、音声指示によって撮像動作を実行する機能を使用するか否かを切り替える音声シャッター切替えスイッチである。
２０３は、回転することにより、デジタルカメラ２００の動作モードを周知の撮影モード、再生モード等に切り替えるモードダイアルである。
２０４は、上下左右の任意方向の指示を入力する四方向選択ボタンである。
２０５は、各種の操作の確定を指示する決定ボタンである。
２０６は、デジタルカメラ２００の電源のＯＮ／ＯＦＦを切り替えるための電源ボタンである。
２０７は、音声入力の開始および終了を指示する手動操作に用いる録音ボタンである。
【００５３】
（音声検出部１０６の説明）
次に、音声検出部１０６の機能の詳細を説明する。
【００５４】
音声検出部１０６は、所定の基準（開始条件）を満たした音を検出し、所定の基準（終了条件）を満たした音を検出する。
【００５５】
続いて所定の基準を満たした音を検出した時点から予め設定された時間が経過した時点で所定の基準を満たした音であることを確定する。
【００５６】
また、入力される音声信号の変化によっては、所定の基準を満たした音ではないと判断する、すなわち所定の基準を満たした音の検出を取り消す。
【００５７】
（音声検出部１０６によって判定される検出状態を示す図）
図３は、音声検出部１０６によって判定された検出状態の一例を示す図である。
【００５８】
音声検出部１０６は、音声信号の検出状況によって仮想的に４つの状態のいずれかに遷移する。
第一状態３０１は、音の入力を開始した直後の状態、すなわち音声信号を検出していない状態（以下、ＳＩＬＥＮＣＥとする）とする。
第二状態３０２は、所定の基準を満たす音の開始を検出し、音の開始の検出を確定していない状態（以下、ＰＯＳＳＩＢＬＥ　ＳＰＥＥＣＨとする）とする。
第三状態３０３は、所定の基準を満たす音の開始が確定した状態（以下、ＳＰＥＥＣＨとする）とする。
第四状態３０４は、音の入力を終了した直後の状態、すなわち音の開始の検出を確定していない状態（以下、ＰＯＳＳＩＢＬＥ　ＳＩＬＥＮＣＥとする）とする。
【００５９】

尚、本実施形態では音の検出状況を仮想的に４つの状態に分類する例を示すが、第二状態３０２と第四状態３０４をまとめて、３つの状態に分類して音の検出状況を判断しても本実施形態と同様の効果が得られる。

【００６０】

（検出状態の遷移についての説明）

第一状態３０１において、音の開始（マイク１１２からの所定の基準を満たす音の入力の開始）を検出すると第二状態３０２に遷移する（３０５）。

第二状態３０２において、音の開始を取り消すと第一状態３０１に遷移する（３０６）。

また、第二状態３０２において、音の開始を確定すると第三状態３０３に遷移する（３０７）。

第三状態３０３において、音の終了（マイク１１２からの所定の基準を満たす音の入力の終了）を検出すると第四状態３０４に遷移する（３０８）。

第四状態３０４において、音の終了を取り消すと第三状態３０３に遷移する（３０９）。

また、第四状態３０４において、所定の基準を満たす音の終了を確定すると音の検出を終了する（３１０）。

第四状態３０４から所定の基準を満たす音の終了を確定すると音の検出を終了させることで、後述する音声認識の処理の際に、音声検出の処理による計算資源、電力等の消費を抑えることが可能となる。

【００６１】

尚、第四状態３０４において所定の基準を満たす音の終了を確定した場合に、第一状態３０１に遷移するようにしてもよい。

第四状態３０４から第一状態３０１に遷移させることで、続けて次の発声を検出することが可能となる。

【００６２】

（音声検出部１０６による処理の概念図）

図４は、音声検出部１０６による処理の一例を示す概念図である。

【００６３】

図４は、ユーザが“Ｓｈｏｏｔ”という言葉を発声した場合の様子を示している。

【００６４】

尚、“Ｓｈｏｏｔ”は撮像を指示するコマンドの一例であり、コマンドの種類については後述する。

【００６５】

図４において４２０は音声信号である。

また、音声信号のうち４２１に示した区間の音声信号はユーザの発声ではなく雑音を検出したものである。

また、音声信号のうち４２２に示した区間の音声信号はユーザが“Ｓｈｏｏｔ”と発声した音を検出したものである。

【００６６】

本実施形態の音声検出部１０６は、所定の基準を満たす音か否かの判断として、音量を検出する。

【００６７】

尚、音量が所定の閾値以上になると発声の開始を検出し、音量が所定の閾値未満になると発声の終了を検出する。

【００６８】

（図４中のパラメータの説明）

図４において、４０１は音声信号４２０から周知の方法で求めた音量（Ｅ（ｔ））、４０２が発声開始を検出するための閾値（ＴＨ１）、４０３が発声終了を検出するための閾値（ＴＨ２）である。

【００６９】

尚、Ｅ（ｔ）は時刻ｔを始点とするフレームにおける音量を表す。

【００７０】
即ち、第一状態３０１でＥ（ｔ）≧ＴＨ１となると発声開始を検出し、第三状態３０３でＥ（ｔ）＜ＴＨ２となると発声終了を検出する。
また、発声開始の検出と発声終了の検出に同じ閾値を用いてもよい（ＴＨ１＝ＴＨ２）。
また、発声開始の検出条件（Ｅ（ｔ）≧ＴＨ１）となるフレームが所定数検出された場合に発声開始を確定する。
【００７１】
同様に、発声終了の検出条件（Ｅ（ｔ）＜ＴＨ２）となるフレームが所定数検出された場合に発声終了を確定する。
【００７２】
本実施形態では、発声開始、発声終了を確定するまでのフレーム数をそれぞれＤ１（例えば、４フレーム）、Ｄ２（例えば、６フレーム）とする。
【００７３】
したがって、第二状態３０２に遷移してからＥ（ｔ）≧ＴＨ１となるフレームがＤ１回検出された場合、発声開始を確定して第三状態３０３に遷移する。
また、第二状態３０２に遷移してからのフレームがＤ１回検出される前に、音量がＥ（ｔ）＜ＴＨ１となった場合、第一状態３０１に遷移する。
【００７４】
尚、第二状態３０２から第一状態３０１に遷移する処理は発声開始を取り消す処理に相当する。
【００７５】
同様に、第四状態３０４に遷移してからＥ（ｔ）＜ＴＨ２となるフレームがＤ２回検出された場合、発声終了を確定し音声検出を終了する。
【００７６】
また、第四状態３０４に遷移してからのフレームがＤ２回検出される前に、音量がＥ（ｔ）≧ＴＨ２となった場合、第三状態３０３に遷移する。
尚、第四状態３０４から第三状態３０３に遷移する処理は発声終了を取り消す処理に相当する。
【００７７】
尚、発声の開始を確定するまでに必要なフレーム数Ｄ１は発声の終了を確定するまでに必要なフレーム数Ｄ２よりも小さ場合いが一般的であるが、同じ数（Ｄ１＝Ｄ２）であってもよい。
【００７８】
４３０は音声信号４２０に対する音声検出部１０６が判定した認識状態の様子を示している。
音声入力開始後は第一状態３０１である。
【００７９】
音量４０１が閾値ＴＨ１以上となる時点ｔ１を始点とするフレームで発声開始を検出（４０４）して第二状態３０２に遷移する。
続いて、時点ｔ２を始点とするフレームでは、第二状態３０２に遷移してからのフレーム数がＤ１回となる前に音量４０１が閾値ＴＨ１未満となるので発声開始を取り消し（４０５）、第一状態３０１に遷移する。
続いて、時点ｔ３を始点とするフレームでは、再び音量４０１が閾値ＴＨ１以上になるので発声開始を検出（４０６）して、第二状態３０２に遷移する。
【００８０】
第二状態３０２に遷移してから音量４０１がＴＨ１以上となるフレーム数がＤ１回となる時点ｔ４で発声開始を確定（４０７）して第三状態３０３に遷移する。
【００８１】
第三状態３０３において、音量４０２が発声終了を検出するための閾値ＴＨ２未満となる時点ｔ５を始点とするフレーム発声終了を検出（４０８）して第四状態３０４に遷移す

る。
続く時点ｔ６を始点とするフレームにおいて音量４０１が閾値ＴＨ２以上となるので発声終了を取り消し（４０９）、第三状態３０３に遷移する。
続く時点ｔ７を始点とするフレームで再び音量４０１が閾値ＴＨ２未満となるので発声終了を検出（４１０）して第四状態３０４に遷移する。
【００８２】
以後、第四状態３０４に遷移してから音量４０１が閾値ＴＨ２未満となるフレームの数がＤ２回となる時点ｔ８で発声終了を確定する（４１１）。
【００８３】
また、発声開始の確定、および発声終了の確定は、フレームの数ではなく、音量が閾値以上あるいは閾値未満である状態が一定の時間継続するか否かで判断してもよい。
【００８４】
すなわち、閾値（ＴＨ１）以上の音量が、発声開始を確定するまでのフレーム数Ｄ１（例えば、４フレーム）に相当する時間Ｓ１（１０２．４ミリ秒間）検出された場合、発声の開始を確定する。
【００８５】
同様に、閾値（ＴＨ２）以下の音量が、発声終了を確定するまでのフレーム数Ｄ２（例えば、６フレーム）に相当する時間Ｓ１（１５３．６ミリ秒間）検出された場合、発声の終了を確定する。
【００８６】
尚、所定の音量が断続して検知された場合でも、継続とみなして継続時間を判断してもよい。
【００８７】
このような構成とすることで、音声検出部１０６は、検出するべき音が一瞬途切れ、対応するフレームの音量が取得できない場合があっても、音が一瞬途切れた後すぐに発声される場合には、適切な処理を実行することが可能となる。
【００８８】
（音声検出部１０６による処理動作を示すフローチャート）
図５は、音声検出部１０６による処理動作を示すフローチャートである。
ステップＳ５０１で、発声開始を検出した時にフレーム番号を初期化する。
【００８９】
以下、フレーム単位で音声の検出を行う。
即ち、音声検出部１０６はフレーム単位に処理を行う際に、当該フレーム毎に音量を計算する。
尚、音量は例えば、対数パワー等信号強度に係る値を周知の方法で音声信号から算出する。
【００９０】
尚、短時間の対数パワーは例えば次式で算出する。
Ｅ（ｔ）＝ｌｏｇ｛Σ（ｘ（ｔ，ｉ）＾２）／Ｎ｝　（１≦ｉ≦Ｎ）・・・（数１）
ここで、Ｎはフレームあたりの音声信号のサンプル数、ｉはフレーム内の音声サンプルのインデックスである。
【００９１】
また、ｘ（ｔ，ｉ）は時点ｔを始点とするフレーム内のｉ番目サンプルの音声信号を表している。
また、ｘ（ｔ，ｉ）＾２はｘ（ｔ，ｉ）の２乗を意味する。
次に、ステップＳ５０２で、第一状態３０１における処理を開始する。
【００９２】
次に、ステップＳ５０３で、時点ｔを始点とするフレームにおける音量Ｅ（ｔ）が発声の開始を検出するために用いる閾値ＴＨ１以上であるか判断する。
音量Ｅ（ｔ）がＴＨ１以上の場合（ステップＳ５０３においてＹＥＳ）、ステップＳ５０

５で第二状態３０２に遷移する。
音量Ｅ（ｔ）がＴＨ１未満の場合（ステップＳ５０３においてＮＯ）、次のフレームの処理（ステップＳ５０４）を繰り返す。
【００９３】
次に、ステップＳ５０６で、第二状態３０２に遷移したフレームを発声開始フレームＴｓと設定する。
【００９４】
次に、ステップＳ５０７で、音量Ｅ（ｔ）がＴＨ１未満であるか判断する。
音量Ｅ（ｔ）がＴＨ１未満の場合（ステップＳ５０７においてＹＥＳ）、第一状態３０１に遷移する。
音量Ｅ（ｔ）がＴＨ１以上の場合（ステップＳ５０７においてＮＯ）、ステップＳ５０８で、第二状態３０２に遷移してからのフレーム数がＤ１回未満であるか判断する。
【００９５】
第二状態３０２に遷移してからのフレーム数がＤ１回未満である場合（ステップＳ５０８においてＹＥＳ）、次のフレームの処理（ステップＳ５０９）を繰り返す。
【００９６】
第二状態３０２に遷移してからのフレーム数がＤ１回以上である場合（ステップＳ５０８においてＮＯ）、ステップＳ５１０で、第三状態３０３に遷移する。
【００９７】
次に、ステップＳ５１２で、音量Ｅ（ｔ）が発声終了の検出に用いる閾値ＴＨ２未満であるか判断する。
音量Ｅ（ｔ）がＴＨ２未満の場合（ステップＳ５１２においてＹＥＳ）、ステップＳ５１４で第四状態３０４に遷移する。
音量Ｅ（ｔ）がＴＨ２以上の場合（ステップＳ５１２においてＮＯ）、ステップＳ５１３で次のフレームの処理を行う。
【００９８】
次に、ステップＳ５１５で、第四状態３０４に遷移したフレームを発声終了フレームＴｅと設定する。
【００９９】
次に、ステップＳ５１６で、音量Ｅ（ｔ）がＴＨ２以上であるか判断する。
音量Ｅ（ｔ）が閾値ＴＨ２以上の場合（ステップＳ５１６においてＹＥＳ）、第三状態３０３に遷移する。
音量Ｅ（ｔ）がＴＨ２未満の場合（ステップＳ５１６においてＮＯ）、ステップＳ５１７で第四状態３０４に遷移してからのフレーム数がＤ２回未満であるか判断する。
【０１００】
第四状態３０４に遷移してからのフレーム数がＤ２回未満である場合（ステップＳ５１７においてＹＥＳ）、ステップＳ５１８で次のフレームの処理を行う。
【０１０１】
第四状態３０４に遷移してからのフレーム数がＤ２回以上である場合（ステップＳ５１７においてＮＯ）、ステップＳ５１９で音声の検出を終了するか判断する。
音声の検出を終了する場合（ステップＳ５１９においてＹＥＳ）、ステップＳ５２０で音声の検出を終了する。
音声検出を終了しない場合（ステップＳ５１９においてＮＯ）、次の発声の検出に備える場合は第一状態３０１に遷移する。
【０１０２】
以上の処理によって音声検出部１０６はフレームＴｓからフレームＴｅまでを発声区間として検出する。
【０１０３】
音声認識部１０７は音声検出部１０６が検出した発声区間（フレームＴｓからフレームＴｅまで）の音声信号を処理して音声認識結果を求める。

【０１０４】
尚、フローチャートと用いた上記の説明では音量の変化に基づいて音声区間を検出する場合を説明したが、これに限定しなくてもよい。
【０１０５】
また、音声検出を行う場合、零交差回数、ピッチ、音声モデルと非音声モデルが出力する尤度比等の周知の特徴量やこれらを組み合わせた特徴量を用いてもよい。
【０１０６】
このような特徴量を用いることで、例えば、周囲から入力される音が大きいような環境下においても発声開始および発声終了を効率良く検出することが可能となる。
【０１０７】
尚、発声開始および発声終了を確定する条件は、以下に示すようにフレーム数以外の条件を用いてもよい。
【０１０８】
例えば、発声開始を検出するための閾値ＴＨ１よりさらに大きな音量である所定の閾値ＴＨ３を設け、発声開始を検出後、音量が所定の閾値ＴＨ３に達したフレームで発声開始を確定してもよい。
【０１０９】
また、発声終了の確定に対しては、発声終了を検出する閾値ＴＨ２よりも小さな音量である所定の閾値ＴＨ４を設けて、発声終了を検出後、音量が所定の閾値ＴＨ４よりも小さくなったフレームで発声終了を確定してもよい。
【０１１０】
このような条件用いて判定することで、発声開始および発声終了を確定するまでの時間を短縮させることが可能となる。
【０１１１】
次に、以上の構成を備えたデジタルカメラ２００において、音声の指示によって撮像動作を実行する場合について説明する。
【０１１２】
（音声検出部１０６、撮像制御部１２３、画像記憶制御部１０４の処理の対応関係）
図１７は、音声検出部１０６、撮像制御部１２３、画像記憶制御部１０４の処理の一例を示す図である。
尚、図３と共通の要素には同一符号を付し、その説明を省略する。
【０１１３】
図１７において、発声開始を検出すると（３０５）、撮像制御部１２３は撮像部１０３に撮像動作を実行させる。
尚、発声開始を検出した場合（３０５）とは、図５のステップＳ５０３のＹＥＳと判断された場合に相当する。
【０１１４】
また、発声終了を検出すると（３０８）、撮像制御部１２３は撮像部１０３に撮像動作を実行させる。
尚、発声終了を検出した場合（３０８）とは、図５のステップＳ５１２のＹＥＳと判断された場合に相当する。
【０１１５】
即ち、撮像部１０３は、音声検出処理の内部状態が第一状態３０１から第二状態３０２に遷移する時点および第三状態３０３から第四状態３０４に遷移する時点で撮像する。
【０１１６】
また、画像記憶制御部１０４は、一旦撮像した画像を、発声開始を取り消した場合（３０６）、発声終了を取り消した場合（３０９）で消去する。
【０１１７】
尚、発声開始を取り消した場合（３０６）とは、図５のステップＳ５０７のＹＥＳと判断された場合である。

【０１１８】
また、発声終了を取り消した場合（３０９）とは、図５のステップＳ５１６のＹＥＳと判断された場合である。
【０１１９】
即ち、図１７において、発声開始を取り消すと、画像制御部１０４は発声開始を検出した場合（３０５）に撮像した画像を消去する。
【０１２０】
同様に、発声終了を取り消すと、画像制御部１０４は発声終了を検出した場合（３０８）に撮像した画像を消去する。
【０１２１】
即ち、第二状態３０２から第一状態３０１に遷移する時、第四状態３０４から第三状態３０３に遷移する時に直前の遷移で撮像した画像を消去する。
【０１２２】
（音声認識文法）
図９は、本実施形態で利用する音声認識文法の一例を示す図である。
【０１２３】
この例では音声認識文法９００はルールを記述する部分９１０と認識するコマンドおよび発音を記述する部分９２０で構成される。
【０１２４】
コマンドおよび発音を記述する部分９２０には一行毎に単語のＩＤ９２１、コマンド９２２、発音９２３が記述されている。
【０１２５】
尚、ルールを記述する部分９１０には、９２２に記載された計９語を認識するための方法が音声認識部１０７に読み取り可能なプログラムコードの形態で記述されている。
【０１２６】
“Ｓｈｏｏｔ”、“Ｇｏ”、“チーズ（Ｃｈｅｅｓｅ）”、“はい、ちーず（Ｓａｙ Ｃｈｅｅｓｅ）”、“Ｆｉｖｅ Ｆｏｕｒ Ｔｈｒｅｅ”は後述する撮像を音声で指示するためのコマンドである。
【０１２７】
“Ｓｐｏｔ Ｍｅｔｅｒｉｎｇ”（スポット測光）、“Ｃｅｎｔｅｒ Ｍｅｔｅｒｉｎｇ”（中央部重点測光）、“Ｕｓｅ ａ ｆｌａｓｈ”（ストロボ発光）、“Ｎｏ Ｆｌａｓｈ”（ストロボ発光禁止）は音声で撮影条件を設定するためのコマンドである。
【０１２８】
以下の説明において、本実施形態のデジタルカメラ２００では、図９に示した音声認識文法９００を言語モデルとして用いる。
【０１２９】
なお本実施例においては、音声を好適な例として説明するが、本発明はこれに限らない。例えば、各音声コマンドの代わりに、何かしらの意味に置換できる音を適用することもできるであろう。例えば、笑い声、列車が通過する際に発生する音等を適用可能である。なお、この場合には、音声認識の技術ではなく、公知の音の種類の検知技術を代用することになるであろう。
【０１３０】
このような構成とすることで、音声に限らず、特徴のある音がマイク１１２を介して入力された場合にも、ユーザは特徴のある各種の音に応答したタイミングで撮像された画像を得ることが可能となる。
【０１３１】
（認識結果制御テーブル）
認識結果制御テーブルは、認識結果に対応する撮像、測光、ストロボ発光等の処理を記述したテーブル形式のデータであり、認識結果処理部１０８が認識結果に対応するカメラ制御を決定する際に参照する。

【０１３２】
尚、認識結果制御テーブルは、認識結果処理部１０８が読み取り可能なプログラムコードの形態でメモリ（認識結果制御テーブル用）１１４に格納されている。
【０１３３】
図１０は、認識結果制御テーブルの一例を示す図である。
図１０において、１０００は認識結果処理データである。
９１０に認識に利用するコマンドが記述されており、１００２に９１０のコマンドに対応するデジタルカメラ２００の制御内容が記述されている。
【０１３４】
（音声によって撮像を指示する場合のデジタルカメラ２００における処理の一例を示すフローチャート）
図６～図８は、音声によって撮像を指示する場合のデジタルカメラ２００における処理の一例を示すフローチャートである。
【０１３５】
まず図６のフローチャートを参照して説明する。
【０１３６】
ステップＳ６０１で、音声シャッター機能がオンに設定されているか否か判断する。
【０１３７】
音声シャッター機能がオンに設定されている場合（ステップＳ６０１においてＹＥＳ）、ステップＳ６０２で録音ボタン２０７を押下されて音声入力を開始する操作が行われたか否か判断する。
【０１３８】
音声シャッター機能がオフに設定されている場合（ステップＳ６０１においてＮＯ）、ステップＳ６９９で音声シャッター機能以外の処理を行う。
【０１３９】
尚、ユーザは操作部１０２が備える音声シャッター切替えスイッチ２０２を操作して、音声シャッター機能のオン・オフを切り替える。
【０１４０】
また、音声シャッター機能のオン・オフの判断は制御部１０１が行う。
【０１４１】
音声入力を開始する操作が行われた場合（ステップＳ６０２のＹＥＳ）、ステップＳ６０３で音声入力部１０５は音声入力の処理を開始し、音声検出部１０６は音声を検出する処理を開始する。
【０１４２】
音声入力を開始する操作以外の操作が行われた場合（ステップＳ６０２のＮＯ）、ステップＳ６９９で音声シャッター機能以外の処理を行う。
【０１４３】
尚、音声入力を開始する操作は、録音ボタン２０７を押下する操作以外の操作でもよい。
【０１４４】
例えば、オートフォーカス機能を備えたデジタルカメラでは、シャッターボタン２０１を半押しすると、焦点を合わせる動作をするものがある。
【０１４５】
この時、オートフォーカス機能の動作に連動して音声入力の処理を開始するようにしてもよい。即ち、ユーザがシャッターボタン２０１を半押しすると音声入力および音声検出の処理を開始する。
【０１４６】
このような構成とすることで、手動作による操作が簡略化される。従って、ユーザは音声入力の処理を素早く開始することができる。
【０１４７】

また、手動作によって音声検出の開始することなく、音声入力部１０５に音声信号が入力された時点で音声検出を開始するようにしてもよい。

【０１４８】

このような構成とすることで、音声検出の処理を素早く開始することができる。また、ユーザがカメラを手動作で操作できない場合にも音声検出を開始すすることができる。従って監視カメラ、防犯カメラ、高所に据え置きされたカメラ等に利用することができる。

【０１４９】

ステップＳ６０４で、音声検出部１０６が発声開始を検出したか否か判断する。

【０１５０】

尚、ステップＳ６０４において発声開始を検出したか否かの判断は、音声検出部１０６が第一状態３０１から第二状態３０２に遷移させる処理を実行したか否かという判断に基づく。

【０１５１】

発声開始を検出した場合（ステップＳ６０４においてＹＥＳ）、ステップＳ６０５で撮像部１０３が撮像動作を実行する。

【０１５２】

ステップＳ６０６で、画像記憶制御部１０４は直前のステップＳ６０５で撮像された画像の第１の画像データをメモリ１１０に記憶させる。

【０１５３】

尚、ステップＳ６０５で撮像した画像、即ち音声検出部１０６が発声開始を検出した時点で撮像した画像を画像Ａとする。

【０１５４】

発声開始を検出しなかった場合（ステップＳ６０４においてＮＯ）、発声開始の検出を繰り返す。

【０１５５】

ステップＳ６０７で、音声検出部１０６が発声開始を取り消すか否か判断する。

【０１５６】

尚、ステップＳ６０５において発声開始を取り消すか否かの判断は、音声検出部１０６が第二状態３０２から第一状態３０１に遷移させる処理を実行したか否かという判断に基づく。

発声開始を取り消す場合（ステップＳ６０７においてＹＥＳ）、ステップＳ６０８で画像記憶制御部１０４はメモリ１１０に記憶した画像Ａを消去する。

発声開始を取り消さない場合（ステップＳ６０７においてＮＯ）、ステップＳ６０９で音声検出部１０６は発声開始を確定したか否か判断する。

【０１５７】

尚、ステップＳ６０９において発声開始を確定したか否かの判断は、音声検出部１０６が第二状態３０２から第三状態３０３に遷移させる処理を実行したか否かという判断に基づく。

発声開始を確定した場合（ステップＳ６０９のＹＥＳ）、ステップＳ６１０で音声認識部１０７が音声認識処理を開始する。

発声開始を確定しなかった場合（ステップＳ６０９のＮＯ）、発声開始を取り消すか否かの判定を繰り返す。

【０１５８】

以降の処理は図７のフローチャートを参照して説明する。

ステップＳ７１１で、音声検出部１０６は発声終了を検出したか否か判断する。

【０１５９】

尚、ステップＳ７１１において発声終了を検出したか否かの判断は、音声検出部１０６が第三状態３０３から第四状態３０４に遷移させる処理を実行したか否かという判断に基づく。

【０１６０】

発声終了を検出した場合（ステップＳ７１１においてＹＥＳ）、ステップＳ７１２で撮像部１０３が撮像を行う。

【０１６１】

次に、ステップＳ７１３で、画像記憶制御部１０４が直前のステップＳ７１２で撮像した画像の第２の画像データをメモリ１１０に記憶する。

尚、ステップＳ７１２で撮像した画像、即ち音声検出部１０６が発声終了を検出した時点で撮像した画像を画像Ｂとする。

尚、一般に「はい、ちーず（Ｓａｙ　Ｃｈｅｅｓｅ）」等の掛け声をかけてに、その発声が終了した後（「ず」を発声した後）に一拍（例えば、０．５秒間）遅延して撮像する場合がある。

【０１６２】

これを考慮して、本実施例では、音声検出部１０６が「はい、ちーず（Ｓａｙ　Ｃｈｅｅｓｅ）」の発声終了を検出した時点から一定の遅延時間が経過してから、撮像部１０３が撮像を行う。

【０１６３】

このような構成とすることで、ユーザが希望する撮像タイミングの種類を増やすことができる。

【０１６４】

次に、ステップＳ７１５で、音声検出部１０６は発声の終了を取り消すか否か判断する。

【０１６５】

尚、ステップＳ７１５において発声終了を取り消すか否かの判断は、音声検出部１０６の第四状態３０４から第三状態３０３に遷移したと認識したか否かという判断に基づく。

【０１６６】

発声終了を取り消す場合（ステップＳ７１５においてＹＥＳ）、ステップＳ７１４で画像記憶制御部１０４はメモリ１１０に記憶された画像Ｂを消去する。

【０１６７】

次に、ステップＳ７１６で、音声検出部１０６が発声終了を確定するか否か判断する。

【０１６８】

尚、ステップＳ７１６において発声終了を確定したか否かの判断は、音声検出部１０６の第四状態３０４から状態遷移を終了したか否かという判断に基づく。

【０１６９】

発声終了を確定した場合（ステップＳ７１６においてＹＥＳ）、ステップＳ７１７で音声入力部１０５および音声検出部１０６による処理を終了する。

【０１７０】

次に、音声検出終了後のステップＳ７１８で、音声認識部１０７は音声検出部１０６が検出した発声区間の音声信号を全て処理するまで音声認識処理を行う。

【０１７１】

音声認識の処理が終了した場合（ステップＳ７１８においてＹＥＳ）、ステップＳ７１９で認識結果処理部１０８は音声認識部１０７が求めた認識結果を取得する。

【０１７２】

以降の処理は図８のフローチャートを参照して説明する。

【０１７３】

ステップＳ８２１で、認識結果処理部１０８は取得した認識結果の認識スコアに基づいて対応するコマンドを受理するか棄却するか判断する。

【０１７４】

尚、コマンドを受理するとは、制御部１０１が認識されたコマンドに対応する制御を決定することをいう。また、コマンドを棄却するとは、制御部１０１が認識されたコマンドに対応する制御が決定されないことをいう。

【０１７５】

取得した認識スコアが所定の閾値以上であり、対応するコマンドを受理した場合（ステップＳ８２１においてＹＥＳ）、ステップＳ８２２で認識結果制御テーブルを参照して認識結果に含まれるコマンドに対応するカメラの制御を決定する。

【０１７６】

認識されたコマンドが発声開始時点で撮像を指示する語（“Ｓｈｏｏｔ”または“Ｇｏ”）の場合（ステップＳ８２２においてＹＥＳ）、ステップＳ８２３で画像記憶制御部１０４がメモリ１１０に記憶されている画像Ａの画像データを記憶媒体１１１に保存する。

【０１７７】

尚、ステップＳ８２３の処理は、認識結果処理部１０８の決定にしたがった処理である。

【０１７８】

次に、ステップＳ８２４で、撮像した画像をユーザが確認できるように表示制御部１０９は画像Ａをディスプレイ１１５に表示する。

認識されたコマンドが発声開始時点で撮像を指示する語（“Ｓｈｏｏｔ”または“Ｇｏ”）でない場合（ステップＳ８２２においてＮＯ）、ステップＳ８２６で、発声終了時点で撮像を指示する語（“チーズ（Ｃｈｅｅｓｅ）”）であるか否か判断する。

認識されたコマンドが発声終了時点で撮像を指示する語（“Ｃｈｅｅｓｅ”）の場合（ステップＳ８２６においてＹＥＳ）、ステップＳ８２７で画像記憶制御部１０４が画像Ｂの画像データを記憶媒体（画像記憶用）１１１に保存する。

【０１７９】

尚、ステップＳ８２７の処理は、認識結果処理部１０８の決定にしたがった処理である。

【０１８０】

ステップＳ８２８で、撮像した画像をユーザが確認できるように表示制御部１０９は画像Ｂをディスプレイ１１５に表示する。

【０１８１】

認識されたコマンドが撮像を指示する語以外の語（“Ｓｐｏｔ　Ｍｅｔｅｒｉｎｇ”等）の場合（ステップＳ８２６においてＮＯ）、ステップＳ８２９で、認識結果処理部１０８が認識結果制御テーブル１１４を参照して、撮像以外のカメラの制御を行う。

【０１８２】

ステップＳ８２５で画像記憶制御部１０４がメモリ１１０に記憶している全ての画像（画像Ａおよび画像Ｂ）の画像データを消去する。

【０１８３】

即ち、所定のコマンドが認識されず、認識結果が棄却されると撮像部１０３が撮像した画像は消去される。

【０１８４】

この処理により、周囲の雑音や認識対象語以外の発声、ユーザ以外の話し声等カメラ操作を意図しない音声の認識結果を棄却し、これらの音を誤って検出して撮像した画像を自動的に消去する。

【０１８５】

尚、判定に用いる閾値は、あらかじめ決めた固定値でもよいし、ガーベッジモデルが出力する認識スコアをｒ倍（０＜ｒ）した値を用いてもよい。

【０１８６】

ガーベッジモデルとは、音声以外の雑音区間や想定される複数の未知語を用いて作成した音響モデルであり、音声認識用データ１１３に含まれる。

【０１８７】

尚、ステップＳ８２２～ステップＳ８２９の処理は、発声開始時点に撮像した画像と発声終了時点で撮像した画像とから認識結果に従って保存する画像を決定している。

【０１８８】

したがって、ユーザは発声内容によって保存する画像の撮像タイミングを自由に変える

ことができる。
【０１８９】
尚、上記の説明では、ステップＳ８２５の後に処理を終了するように説明したが、引き続き次の音声入力を行うため、ステップＳ６０２の処理へ進んでも良い。
【０１９０】
このように構成し、シャッターボタン２０１を半押しすることで音声入力開始を操作する場合は、シャッターボタン２０１を半押ししている間は何度でも音声入力によるカメラ制御が可能になる。
【０１９１】
例えば、シャッターボタン２０１を半押ししたまま、“Ｃｅｎｔｅｒ Ｍｅｔｅｒｉｎｇ”等の発声で撮影条件を設定し、つづく発声で撮像指示を出す、といったことが可能になる。
【０１９２】
（“Ｓｈｏｏｔ”で撮像する場合の説明）
図１１は本実施形態に係るデジタルカメラ２００を利用して、“Ｓｈｏｏｔ”という音声指示で撮像する場合の動作示す図である。
【０１９３】
図１１の横軸１１５０は時間軸であり、左から右に時刻が推移する。ｔ１～ｔ７は時点を示している。
１１１０は音声入力部１０５がＡ／Ｄ変換した音声信号である。
１１１１はユーザが“Ｓｈｏｏｔ”と発声した区間の音声信号（音声波形）である。
１１２０は音声信号１１１０に対応する音量の変化を示す。
１１２１は音声検出部１０６で用いる発声開始検出用の閾値（ＴＨ１）、１１２２は発声終了検出用の閾値（ＴＨ２）である。
１１３０は、音声検出部１０６が認識した状態の変化を視覚的に示したものである。
１１４０はデジタルカメラ２００の動作の内容を示している。
【０１９４】
続いて、時点ｔ１から時点ｔ７までの時間経過に沿ってデジタルカメラ２００の動作を説明する。
【０１９５】
（時点ｔ１）
音量１１２０が閾値ＴＨ１以上になる時点ｔ１を始点とするフレームで音声検出部１０６が発声開始を検出する。これは、上述した所定の基準（開始条件）を満たした音を検出する工程に相当する。
この時、音声検出部１０６は第一状態３０１から第二状態３０２に遷移させる処理を実行する（１１３０の時点ｔ１の部分）。
発声開始を検出した時点で撮像部１０３が同時点の被写体（ＩＭＧ００３）を撮像し、続いて、画像記憶制御部１０４が撮像した画像の画像データをメモリ（画像記憶用）１１０に記憶する（以上、１１４１）。
【０１９６】
（時点ｔ２）
音声検出部１０６は、発声開始を検出した時点ｔ１を始点とするフレームからＤ１番目のフレームである時点ｔ２を始点とするフレームで発声開始を確定する。
同時に、音声認識部１０７による音声認識の処理を開始する（以上１１４２）。
この時、音声検出部１０６は、第二状態３０２から第三状態３０３に遷移させる処理を実行する（１１３０の時点ｔ２の部分）。
【０１９７】
（時点ｔ３）
続いて、音量１１２０が閾値ＴＨ２未満になる時点ｔ３を始点とするフレームで音声検出部１０６が発声終了を検出する。これは、上述した所定の基準（終了条件）を満たした

音を検出する。
この時、音声検出部１０６は、第三状態３０３から第四状態３０４へ遷移させる処理を実行する（１１３０の時点ｔ３の部分）。
音声検出部１０６が発声終了を検出した時点ｔ３で、撮像部１０３がこの時点の被写体（ＩＭＧ００５）を撮像し、続いて画像記憶制御部１０４が撮像した画像を撮像した画像の画像データをメモリ（画像記憶用）１１０に記憶する（以上、１１４３）。
【０１９８】
（時点ｔ４）
音声検出部１０６が発声終了を検出した時点ｔ３を始点とするフレームからＤ２番目のフレームとなる前の、時点ｔ４を始点とするフレームで音量１１２０が閾値ＴＨ２以上になると、音声検出部１０６は発声終了を取り消す。
この時、音声検出部１０６は第四状態３０４から第三状態３０３に遷移させる処理を実行する（１１３０の時点ｔ４の部分）。
発声終了が取り消された時点ｔ４で、画像記憶制御部１０４は発声終了を検出した時点ｔ３で撮像した画像ＩＭＧ００５の画像データをメモリ（画像記憶用）１１０から消去する（以上、１１４４）。
【０１９９】
（時点ｔ５）
続く時点ｔ５を始点とするフレームで音量１１２０が閾値ＴＨ２未満になるので、音声検出部１０６が発声終了を検出する。
この時、音声検出部１０６は第三状態３０３から第四状態３０４に遷移させる処理を実行する（１１３０の時点ｔ５の部分）。
また、撮像部１０３はこの時点ｔ５の被写体（ＩＭＧ００６）を撮像し、画像記憶制御部１０４が撮像した画像の画像データをメモリ（画像記憶用）１１０に記憶する（以上１１４５）。
【０２００】
（時点ｔ６）
発声終了を検出した時点ｔ５を始点とするフレームから、音量１１２０が閾値ＴＨ２以上になることなくＤ２番目のフレームである時点ｔ６を始点とするフレームで、音声検出部１０６は発声の終了を確定する（１１４６）。
この時、前述したように、音声検出部１０６は第四状態３０４から第１状態に遷移させる処理を実行してもよく、音声検出部１０６は状態を遷移させる処理を終了させてもよい。
【０２０１】
（時点ｔ７）
その後、音声認識部１０７の処理が終了した時点ｔ７で認識結果処理部１０８がデジタルカメラ２００の制御方法を決定する。
ここで“Ｓｈｏｏｔ”が認識結果として得られた場合は、認識結果制御テーブルを参照して“Ｓｈｏｏｔ”に対応する処理を決定する。
図１０に示したように“Ｓｈｏｏｔ”は発声の開始を検出した時点での撮像動作と対応付けられたコマンドである。
認識結果処理部１０８の決定にしたがって、画像記憶制御部１０４が発声開始を検出した時点ｔ１で撮像した画像（ＩＭＧ００３）の画像データを画像記憶媒体１１１に保存する。
同時に、画像記憶制御部１０４は発声終了時点で撮像した画像（ＩＭＧ００６）を保存せずにメモリ（画像記憶用）１１０から消去する。
【０２０２】
（“チーズ（Ｃｈｅｅｓｅ）”で撮像する場合の説明）
図１２は、本実施形態に係るデジタルカメラ２００を利用して、“チーズ（Ｃｈｅｅｓｅ）”という音声指示で撮像する場合の動作を示す図である。
【０２０３】

図１１と同様に１２５０は時間軸であり、１２１０は音声信号、１２２０は音量、１２３０は音声検出部１５０が認識した状態、１２４０はデジタルカメラ２００の動作を示す。
１２１１はユーザの発声前に混入した雑音を検知した区間であり、１２１２はユーザが発した“Ｃｈｅｅｓｅ”という音声を検知した区間である。
１２２１は音声検出部１０６で用いる発声区間を検出するための閾値（ＴＨ１）である。

【０２０４】

尚、図１２では、発声開始、発声終了の検出に同じ閾値ＴＨ１を用いる。

【０２０５】

以下、時間経過に沿ってデジタルカメラ２００の動作を説明する。

【０２０６】

（時点ｔ１）
時点ｔ１を始点とするフレームで音声検出部１０６が発声開始を検出すると、撮像部１０３が時点ｔ１を始点とするフレームに対応する被写体１２０２（ＩＭＧ００２）を撮像する。また、画像記憶制御部１０４が撮像した画像の画像データをメモリ（画像記憶用）１１０に一時的に記憶する（１２４１）。

【０２０７】

（時点ｔ２）
時点ｔ２を始点とするフレームで、発声開始を検出してからのフレーム数がＤ１回となる前に音量が閾値ＴＨ１未満になるため、音声検出部１０６が発声開始を取り消す。
この時、画像記憶制御部１０４が１２４１で撮像したＩＭＧ００２を消去する。

【０２０８】

（時点ｔ３）
時点ｔ３を始点とするフレームで音声検出部１０６が再び発声開始を検出すると、撮像部１０３が時点ｔ３を始点とするフレームに対応する被写体１２０３（ＩＭＧ００３）を撮像する。また、画像記憶制御部１０４が撮像した画像の画像データをメモリ（画像記憶用）１１０に一時的に記憶する（１２４３）。

【０２０９】

（時点ｔ４）
時点ｔ４を始点とするフレームで音声検出部１０６が発声開始を確定すると、音声認識部１０７が音声認識の処理を開始する（１２４４）。

【０２１０】

（時点ｔ５）
時点ｔ５を始点とするフレームで音声検出部１５０が発声終了を検出すると、撮像部１０３が時点ｔ５始点とするフレームに対応する被写体の画像１２０５（ＩＭＧ００５）を撮像する。また、続いて画像記憶制御部１０４が撮像した画像の画像データをメモリ（画像記憶用）１１０に一時的に記憶する（１２４５）。

【０２１１】

（時点ｔ６）
時点ｔ６を始点とするフレームで音声検出部１０６が発声終了を確定する（１２４６）。

【０２１２】

（時点ｔ７）
発声終了確定後、音声認識部１０７による音声認識処理が終了する時点ｔ７で、認識結果処理部１０８が得られた認識結果に基づいてカメラの制御を決定する。
尚、図１０に示したように“Ｃｈｅｅｓｅ”は発声の終了を検知した時点での撮像動作と対応付けられたコマンドである。
したがって、画像記憶制御部１０４は、発声終了を検出した時点ｔ５で撮像した画像（ＩＭＧ００５）の画像データを画像記憶媒体１１１に保存し、発声開始を検出した時点ｔ３で撮像した画像（ＩＭＧ００３）の画像データは保存せずに消去する。

【０２１３】
以上、図１１、図１２を用いて説明したように、本実施形態で説明したデジタルカメラ２００では、発声を開始したタイミングの画像を得たい場合は“Ｓｈｏｏｔ”（または“Ｇｏ”）と発声すればよい。
【０２１４】
また、本実施形態で説明したデジタルカメラ２００では、発声を完了したタイミングの画像を得たい場合は“Ｃｈｅｅｓｅ”と発声すればよい。
【０２１５】
また、発声を開始したタイミングから一定時間後（すなわち、“Ｔｗｏ　Ｏｎｅ　Ｚｅｒｏ”が発声されるべき時間後）の時間関係にあるタイミングで撮像された画像の画像データを得たい場合は“Ｆｉｖｅ　Ｆｏｕｒ　Ｔｈｒｅｅ”と発声すればよい。
【０２１６】
また、発声を終了したタイミングから一定時間後（例えば、０．５秒後）の画像の画像データを得たい場合は“はい、チーズ（Ｓａｙ　Ｃｈｅｅｓｅ）”と発声すればよい。
【０２１７】
“Ｓｈｏｏｔ”（または“Ｇｏ”）と発声した場合、音声認識の終了を待たずに撮像するため、乗り物等の動いている被写体を撮影する場合に好適である。
また、“Ｃｈｅｅｓｅ”（または、“Ｓａｙ　Ｃｈｅｅｓｅ”）と発声した場合、発声後に撮影するため、集合写真や記念写真等、被写体に撮影タイミングを伝えて撮影する場合に好適である。
また“Ｆｉｖｅ　Ｆｏｕｒ　Ｔｈｒｅｅ”と発声した場合、発声を開始してから一定時間後（すなわち、“Ｔｗｏ　Ｏｎｅ　Ｚｅｒｏ”が発声されるべき時間後）の所望のタイミングで撮像された画像を得ることができる。
【０２１８】
従って、撮影シーンによって自由に撮影タイミングを変えた撮影が可能になり、ユーザの利便性が向上する。
【０２１９】
また、撮影後にユーザが意図しないタイミングで撮影した画像を手動作によって削除する手間も必要ないという利点がある。
【０２２０】
即ち、図１２で説明したように、音声入力時に混入した周囲の雑音を発声と誤って撮像した場合でも、発声開始が確定されなければ自動的に消去する。
【０２２１】
また、雑音や撮像を意図しない発声により撮像した場合でも、図８のＳ８２１における処理で、撮像を意図しない語が認識された場合は認識結果を棄却して誤って撮像した画像を消去する。
【０２２２】
したがって、音声指示による撮影を行う場合に、周囲雑音による誤動作の影響を少なくするという効果がある。
【０２２３】
（第１の実施形態の変形例１）
本実施形態においては、発声の開始を検出したタイミングさ撮像するか、発声の終了を検出したタイミングで撮像するようにしてもよい。
【０２２４】
（フローチャート）
図１３に発声開始を検出した時点でのみ撮像する場合のフローチャートを示す。
図１３に示したフローチャートは図６～図８で説明した処理と異なるフローチャートになるステップＳ８１１以降の処理のみ示している。
【０２２５】
また、図７～図８と同じ処理については同じ符号で示している。以下、図７～図８と図

３の相違点のみ説明する。
図１３のフローチャートでは、図７のフローチャートで行っていた、発声終了を検出した時に撮像する処理（ステップＳ７１２、ステップＳ７１３）および撮像した画像を消去する処理（ステップＳ７１４）は行わない。
また図１３のフローチャートでは、図８のフローチャートで行っていた、発声終了時に撮像を指示する語を認識した場合の認識結果処理部１０８が行う処理（ステップＳ８２６，ステップＳ８２７、ステップＳ８２８）を行わない。
【０２２６】
その他の処理については、図６～図８で説明した処理と同じである。
【０２２７】
なお、発声開始を検出した時点でのみ撮像する場合は、図９に示した音声認識文法から発声開始時点に撮影を指示する語（“Ｃｈｅｅｓｅ”、“Ｓａｙ　Ｃｈｅｅｓｅ”等）を削除する。
【０２２８】
音声認識文法を変更しない場合は、図１０に示した認識結果制御データを変更し、“Ｃｈｅｅｓｅ”、“Ｓａｙ　Ｃｈｅｅｓｅ”等を認識した場合の処理を、発声の開始を検出した時点で撮像する処理に変更する。
【０２２９】
これにより、ユーザが“Ｃｈｅｅｓｅ”、“Ｓａｙ　Ｃｈｅｅｓｅ”と発声すると、発声開始時点で撮像した画像の画像データが画像記憶媒体１１１に記憶される。
【０２３０】
同様にして、発声終了を検出した時点でのみ撮像するように変形することもできる。この場合は、発声開始を検出したときに撮像する処理（ステップＳ６０５、ステップＳ６０６）および発声開始が取り消されたときの処理（ステップＳ６０８）が省かれる。
【０２３１】
また、認識結果処理部１０８が行う処理のうち、ステップＳ８２２～ステップＳ８２４が省かれる。
【０２３２】
このとき、ステップＳ８２１で認識結果を受理した場合（ステップＳ８２１においてＹＥＳ）に、ステップＳ８２６以降の処理を行う。
【０２３３】
また、発声開始時に撮影を指示する語を音声認識文法９００から削除するか、認識結果制御データに記述された処理内容を変更する。
【０２３４】
（第１の実施形態の変形例２）
本実施形態では、認識結果によって発声の開始を検出した時点および発声の終了を検出した時点の画像の画像データを記憶媒体（画像記憶用）１１１に記憶する構成してもよい。
【０２３５】
たとえば、“Ｓａｙ　Ｃｈｅｅｓｅ”に対して発声開始を検出した時点および発声終了を検出した時点の両方で撮像するように認識結果制御データに記述すれば、両方時点における画像の画像データが記憶媒体（画像記憶用）１１１に記憶される。
【０２３６】
このような構成とすることで、ユーザが指示することが可能な撮像タイミングの種類が増えて、ユーザの利便性が向上する。
【０２３７】
（第１の実施形態の変形例３）
本実施形態では、認識結果処理部１０８が行う処理において、認識結果を棄却した場合（ステップＳ８２１においてＮＯ）、メモリ（画像記憶用）１１０に記憶した画像Ａ、画像Ｂを消去（ステップＳ８２５）するか否かユーザに確認させるようにしてもよい。

【０２３８】
　また、ユーザが記憶媒体（画像記憶用）１１１に記憶する画像を選択するようにしてもよい。
【０２３９】
　また、認識結果が棄却された場合は、画像Ａ、画像Ｂの両方の画像データを記憶媒体（画像記憶用）１１１に記憶するようにしてもよい。
【０２４０】
　例えば、ディスプレイ１１５に画像Ａ、画像Ｂを表示し、画像データの消去の可否を四方向ボタン２０４で選べるようする。
【０２４１】
　また、四方向ボタン２０４で記憶する画像をユーザが選択し、決定ボタン２０５が押された時点で選択されている画像の画像データを記憶媒体（画像記憶用）１１１に記憶するようにする。
【０２４２】
　撮影を指示する語以外が認識された場合（ステップＳ８２６においてＮＯ）についても同様に、画像消去の確認、記憶媒体（画像記憶用）１１１に記憶する画像の選択をユーザが行えるように構成できる。
【０２４３】
　また、画像Ａ、画像Ｂの画像データをともに記憶媒体（画像記憶用）１１１に記憶するようにしてもよい。
【０２４４】
　このように構成することで、音声認識の性能が劣化するような環境で音声指示による撮像機能を使用する場合に、音声認識の誤りによって所望の撮像画像を誤って消去することが防止でき、ユーザの利便性が向上する。
【０２４５】
　尚、メモリ（画像記憶用）１１０の記憶容量に応じて、１回の音声認識において保持する画像の数を決定してもよい。
【０２４６】
　このように構成することで、メモリ１１０が限られた記憶容量を考慮して、ユーザが希望する画像の候補をできるだけ一時的に保持しておくことができる。
【０２４７】
　（第１の実施形態の変形例４）
　本実施形態の認識結果処理部１７０の処理において、撮像タイミングの異なる語の認識スコアの差が所定の閾値より小さい場合は、発声開始時点および発声終了時点で撮像された画像の両方を記憶媒体（画像記憶用）１１１に記憶するようにしてもよい。
【０２４８】
　例えば、発声開始時点での撮像を指示する“Ｓｈｏｏｔ”と発声終了時点での撮像を指示する“Ｃｈｅｅｓｅ”の認識スコアの差が所定値未満の場合に、発声開始時点および発声終了時点で撮像された画像を両方とも記憶媒体（画像記憶用）１１１に記憶する。
【０２４９】
　あるいは、二つの画像をディスプレイ１１５に表示して、ユーザが選択するようにしてもよい。
【０２５０】
　このように構成することで、音声認識の性能が劣化するような環境で音声指示によって撮像を実行する機能を使用する場合に、音声認識の認識誤りによって、所望の撮像画像を誤って消去することが防止でき、ユーザの利便性が向上する。
【０２５１】
　（第１の実施形態の変形例５）
　本実施形態では、撮像した画像の画像データをメモリ（画像記憶用）１１０に一時的に記憶し、認識結果確定後に記憶媒体（画像記憶用）１１１に記憶するよう説明したが、最

初から記憶媒体（画像記憶用）１１１に記憶するようにしてもよい。
【０２５２】
この場合、ステップＳ６０８，ステップＳ７１４における画像データの消去の処理は、記憶媒体（画像記憶用）１１１に記憶された画像データを消去することになる。
【０２５３】
また、ステップＳ８２３、ステップＳ８２７の処理は行わない。
【０２５４】
さらに、認識結果を棄却した場合（ステップＳ８２１においてＮＯ）および認識結果が撮影を指示する語でない場合（ステップＳ８２６においてＮＯ）は、記憶媒体（画像記憶用）１１１に記憶されている画像Ａおよび画像Ｂの画像データを消去する。
【０２５５】
さらに、認識結果が発声開始時点で撮像を指示する語である場合には画像Ｂの画像データを消去し、発声終了時点で撮像を指示する語である場合には画像Ａの画像データを消去する。
【０２５６】
（第１の実施形態の変形例６）
例えば、道路脇等、周囲の雑音の影響を受け易い場所で本実施形態に係るデジタルカメラ２００を使用する場合、音声検出部１０６の内部状態が短時間に頻繁に変化する場合がある。
【０２５７】
短時間のうちに撮像と画像データの消去が繰り返されると、デジタルカメラ２００の連写機能が画像データを消去した直後の撮像に対応しきれず、メモリ（画像記憶用）１１０上に画像が記憶されないということが起こりうる。
【０２５８】
これに対処するため、例えば、ステップＳ１０８で発声開始を検出した時点では撮像した画像Ａの画像データを消去せずに、次の発声開始を検出するまで画像Ａの画像データをメモリ（画像記憶用）１１０に記憶しておいてもよい。
【０２５９】
この場合、次に発声開始を検出した時点で画像Ａの画像データを消去するか、画像Ａの画像データに新たに撮像された画像の画像データを上書きするようにする。
【０２６０】
同様に、ステップＳ７１５において発声終了を取り消した場合も、画像Ｂの画像データを消去せずに、次に発声終了を検出するまでメモリ（画像記憶用）１１０に記憶しておいてもよい。
【０２６１】
このように構成することで、カメラの連写が音声検出の状態変化の速度に間に合わない場合でも、少なくとも最初に撮像された画像は残しておくことができる。
【０２６２】
尚、上記各実施形態では、カメラについて説明したが、本発明はビデオカメラ等の他の撮像装置にも適用することができる。
【０２６３】
（第１の実施形態の変形例７）
本実施形態では、マイク１１２として、周知ステレオマイクを用いる。
【０２６４】
また、音声認識部１０７は、左右のマイク１１２を介して入力されるそれぞれの音声信号の音量、ピッチ等の関係を前述した特徴量として用いてもよい。
【０２６５】
このような特徴量を用いることで、例えば、デジタルカメラ２００に対して右側から迫る音源と左側から迫る音源を判別することができる。すなわち、撮像する際の状況を認識して撮像することが可能となる。

【０２６６】
（第１の実施形態の変形例８）
本実施形態では、認識結果制御テーブルに含まれるコマンドの一例として示した“チーズ（Ｃｈｅｅｓｅ）”に換えて“ハイ、チーズ（Ｓａｙ　Ｃｈｅｅｓｅ）”というコマンドに発声終了時点で撮像する処理を対応付けてもよい。
【０２６７】
また、認識結果制御テーブルに含まれるコマンドの一例として示した“Ｇｏ”に換えて“今”というコマンドに発声開始時点で撮像する処理を対応付けてもよい。
【０２６８】
（第２の実施形態）
図１６は、本発明の第２の実施形態に係る情報処理装置１６００の構成の一例を示す機能ブロック図である。
【０２６９】
尚、図１と共通の要素には同一符号を付し、その説明を省略する。
【０２７０】
情報処理装置１６００は、入力装置１６０２、撮像装置１６０３、格納装置（画像記憶用）１６１０、記憶装置（画像記憶用）１６１１、集音装置１６１２と接続可能であることを特徴とする。
【０２７１】
また、情報処理装置１６００は、格納装置（音声認識データ用）１６１３、格納装置（認識結果制御データ用）１６１４、表示制御装置１６０９と接続可能であることを特徴とする。
【０２７２】
尚、入力装置１６０２は操作部１０２に、撮像装置１６０３は撮像部１０３に、格納装置（画像記憶用）１６１０はメモリ（画像記憶用）１１０に、記憶装置（画像記憶用）１６１１は記憶媒体（画像記憶用）１１１に対応する機能を備える。
【０２７３】
また、集音装置１６１２はマイク１１２に、格納装置（音声認識データ用）１６１３はメモリ（音声認識データ用）１１３に対応する機能を備える。
【０２７４】
また、格納装置（認識結果制御データ用）１６１４は認識結果データ用メモリ１１４に、表示制御装置１６０９は表示制御部１０９に対応する機能を備える。
【０２７５】
情報処理装置１６００としては、例えばマイクロプロセッサ等が想定できる。
【０２７６】
図１４、図１５は、情報処理装置１６００における処理動作の一例を示したフローチャートである。
【０２７７】
まず、図１４のフローチャートを参照して説明する。
ステップＳ１４００で、音声入力部１０５は音声信号が入力されたか否か判断する。
【０２７８】
音声信号が入力された場合（ステップＳ１４００においてＹＥＳ）、ステップＳ１４０１で音声検出部１０６はフレームｆを初期化する（ｆ＝０）。
【０２７９】
次に、ステップＳ１４０２で音声検出部１０６は音声信号の検出状態を第一状態３０１に設定する。
【０２８０】
次に、ステップＳ１４０３で音声検出部１０６は検出の対象となるフレームを設定する。
【０２８１】

次に、ステップＳ１４０４で音声検出部１０６は音声入力部１０５に入力された音声信号の特徴量データを記憶する。

【０２８２】

尚、特徴量データとは、音声認識部１０７が音声認識を行う場合に使用するデータである。

【０２８３】

次に、ステップＳ１４０５で音声検出部１０６は音声の検出状態が第１の状態から第４の状態のいずれであるか判断する。

【０２８４】

ステップＳ１４０５で音声検出部１０６が検出状態を第一状態３０１であると判断した場合、ステップＳ１４０６で、第１の検出として、音声検出部１０６は閾値ＴＨ１以上の音量を検出したか否か判断する。

【０２８５】

閾値ＴＨ１以上の音量を検出した場合（ステップＳ１４０６においてＹＥＳ）、ステップＳ１４０７で、音声検出部１０６は検出状態を第二状態３０２に遷移させる（このタイミングを第１の時刻とする）。

【０２８６】

次に、ステップＳ１４０８で、撮像制御部１２３は撮像装置１６０３に撮像動作を実行させる信号を出力する。

【０２８７】

尚、ステップＳ１４０８で出力された信号によって撮像された画像を画像Ａとする。

【０２８８】

次に、ステップＳ１４０９で、画像記憶制御部１０４は、第１の取得として、直前のステップＳ１４０８によって撮像された画像Ａを表す画像データを格納装置（画像記憶用）１６１０に記憶させる信号を出力する。

【０２８９】

次に、ステップＳ１４１０で、第１の記憶として、音声検出部１０６は処理中のフレームｆを発声開始フレームＦｓとして記憶する。

【０２９０】

次に、ステップＳ１４０３に戻り、音声検出部１０６は次の音声検出の対象となるフレームを設定する。

【０２９１】

また、ステップＳ１４０６で閾値ＴＨ１以上の音量を検出しなかった場合（ステップＳ１４０６においてＮＯ）、同様にステップＳ１４０３に戻り、音声検出部１０６は次の音声検出の対象となるフレームを設定する。

【０２９２】

また、ステップＳ１４０５で音声検出部１０６が検出状態を第二状態３０２であると判断した場合、ステップＳ１４１１で、処理中のフレームｆが発声開始フレームＦｓからＭ１回目のフレーム以上であるか否か判断する。

【０２９３】

また、処理中のフレームｆが発声開始フレームＦｓからＭ１回目のフレーム未満である場合（ステップＳ１４１１においてＹＥＳ）、ステップＳ１４１３で音声検出部１０６が閾値ＴＨ１より大きい音量を検出したか否か判断する。

【０２９４】

閾値ＴＨ１より大きい音量を検出しなかった場合（ステップＳ１４１３においてＮＯ）、ステップＳ１４１４で音声検出部１０６はカウンタＦａの値を初期化する。

【０２９５】

次に、ステップＳ１４０３に戻り、音声検出部１０６は次の音声検出の対象となるフレームを設定する。

【０２９６】

尚、カウンタＦａとは、発声開始フレームＦｓを設定し直すか否か判定するために使用する。

【０２９７】

また、閾値ＴＨ１より未満の音量を検出した場合（ステップＳ１４１３においてＹＥＳ）、ステップＳ１４１５で音声検出部１０６はカウンタＦａの値を１増やす。

【０２９８】

次に、ステップＳ１４１６で音声検出部１０６はカウンタＦａの値がＮ１以上であるか判断する。

【０２９９】

カウンタＦａの値がＮ１以上である場合（ステップＳ１４１６においてＹＥＳ）、ステップＳ１４１７で画像記憶制御部１０４は格納装置（画像記憶用）１６１０に記憶された画像Ａを表す画像データを消去するための信号を出力する。

【０３００】

尚、ステップＳ１４１７における処理は、音声認識後に画像データを消去する処理に対して、第２の消去に相当する。

【０３０１】

次に、ステップＳ１４１８で、発声の開始を再検出する第１の再検出をおこなうために、音声検出部１０６は検出状態を第一状態３０１に遷移させる。

【０３０２】

次に、ステップＳ１４０３に戻り、音声検出部１０６は次の音声検出の対象となるフレームを設定する。

【０３０３】

また、カウンタＦａの値がＮ１未満である場合（ステップＳ１４１６においてＮＯ）、同様にステップＳ１４０３に戻り、音声検出部１０６は次の音声検出の対象となるフレームを設定する。

【０３０４】

また、ステップＳ１４１１で処理中のフレームｆが発声開始フレームＦｓからＭ１回目のフレーム以上である場合（ステップＳ１４１１においてＮＯ）、ステップＳ１４１２で音声検出部１０６は検出状態を第三状態３０３に遷移させる。

【０３０５】

また、ステップＳ１４０５で音声検出部１０６が検出状態を第三状態３０３であると判断した場合、ステップＳ１４１９で、第２の検出として、音声検出部１０６は閾値ＴＨ２以下の音量を検出したか否か判断する。

【０３０６】

閾値ＴＨ２以下の音量を検出した場合（ステップＳ１４１９においてＹＥＳ）、ステップＳ１４２０で、音声検出部１０６は検出状態を第四状態３０４に遷移させる（このタイミングを第２の時刻とする）。

【０３０７】

次に、ステップＳ１４２１で、撮像制御部１２３は撮像装置１６０３に撮像動作を実行させるための信号を出力する。

【０３０８】

尚、ステップＳ１４２１で出力された信号によって撮像された画像を画像Ｂとする。

【０３０９】

次に、ステップＳ１４２２で、画像記憶制御部１０４は、第２の取得として、直前のステップＳ１４２１で撮像された画像Ｂを表す画像データを格納装置（画像記憶用）１６１０に記憶させる信号を出力する。

【０３１０】

次に、ステップＳ１４２３で、第２の記憶として、音声検出部１０６は処理中のフレームｆを発声終了フレームＦｅとして設定する。

【０３１１】

次に、ステップＳ１４０３に戻り、音声検出部１０６は次の音声検出の対象となるフレームを設定する。

【０３１２】

また、ステップＳ１４１９で閾値ＴＨ１以上の音量を検出なかった場合（ステップＳ１４１９においてＮＯ）、同様にステップＳ１４０３に戻り、音声検出部１０６は次の音声検出の対象となるフレームを設定する。

【０３１３】

また、ステップＳ１４０５で音声検出部１０６が検出状態を第四状態３０４であると判断した場合、ステップＳ１４２４で、処理中のフレームｆが発声終了フレームＦｅからＭ２回目のフレーム以上であるか否か判断する。

【０３１４】

また、処理中のフレームｆが発声終了フレームＦｅからＭ２回目のフレーム未満である場合（ステップＳ１４２４においてＹＥＳ）、ステップＳ１４２６で音声検出部１０６が閾値ＴＨ２より大きい音量を検出したか否か判断する。

【０３１５】

閾値ＴＨ２より大きい音量を検出しなかった場合（ステップＳ１４２６においてＮＯ）、ステップＳ１４２７で音声検出部１０６はカウンタＦｂの値を初期化する。

【０３１６】

次に、ステップＳ１４０３に戻り、音声検出部１０６は音声検出の対象となるフレームを設定する。

【０３１７】

尚、カウンタＦｂとは、発声終了フレームＦｅを設定し直すか否か判定するために使用する。

【０３１８】

また、閾値ＴＨ２より大きい音量を検出した場合（ステップＳ１４２６においてＹＥＳ）、ステップＳ１４２８で音声検出部１０６はカウンタＦｂの値を１増やす。

【０３１９】

次に、ステップＳ１４２９で音声検出部１０６はカウンタＦｂの値がＮ２以上であるか判断する。

【０３２０】

カウンタＦｂの値がＮ２以上である場合（ステップＳ１４２９においてＹＥＳ）、ステップＳ１４３０で画像記憶制御部１０４は格納装置（画像記憶用）１６１０に記憶された画像Ｂを表す画像データを消去するための信号を出力する。

【０３２１】

尚、ステップＳ１４３０における処理は、音声認識後に画像データを消去する処理に対して、第３の消去に相当する。

【０３２２】

次に、ステップＳ１４３１で音声検出部１０６は、発声の終了を再検出する第２の再検出をおこなうために、検出状態を第三状態３０３に遷移させる。

【０３２３】

次に、ステップＳ１４０３に戻り、音声検出部１０６は次の音声検出の対象となるフレームを設定する。

【０３２４】

また、カウンタＦｂの値がＮ２未満である場合（ステップＳ１４２９においてＮＯ）、同様にステップＳ１４０３に戻り、音声検出部１０６は次の音声検出の対象となるフレームを設定する。

【０３２５】

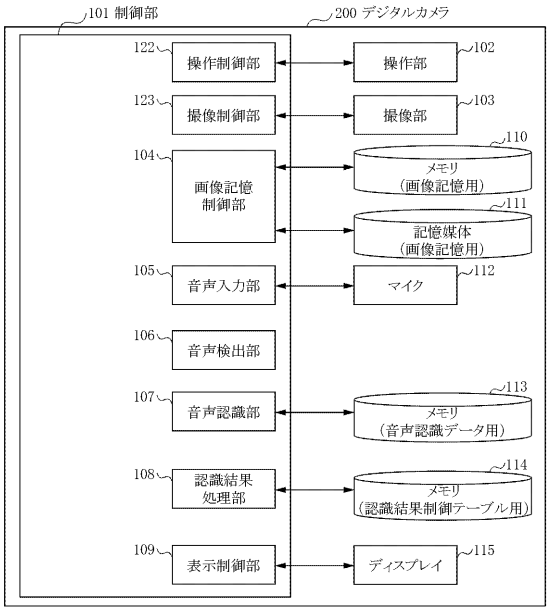
また、ステップＳ１４２４で処理中のフレームｆが発声開始フレームＦｅからＭ２回目のフレーム以上である場合（ステップＳ１４２４においてＮＯ）、ステップＳ１４２５で音声検出部１０６は音声検出を終了する。

【０３２６】
次に、図１５のフローチャートを参照して説明する。
【０３２７】
ステップＳ１５３２で音声認識部１０７はステップＳ１５０４で取得した各フレームの特徴量データと音声認識用データとに基づいて音声認識を行う。
【０３２８】
次に、ステップＳ１５３３で音声認識部１０７による音声認識を終了する。
【０３２９】
尚、ステップＳ１５３３の処理は、音声認識部１０７によって音声認識の結果が得られた後に実行する。
【０３３０】
次に、ステップＳ１５３４で、認識結果処理部１０８は音声認識の結果が発声開始のタイミングで撮像を指示する内容であるか否か判断する。
【０３３１】
発声開始のタイミングで撮像を指示する内容である場合（ステップＳ１５３４のＹＥＳ）、ステップＳ１５３５で画像Ｂを消去するための信号を出力する。
【０３３２】
発声開始のタイミングで撮像を指示する内容でない場合（ステップＳ１５３４のＮＯ）、ステップＳ１５３６で、認識結果処理部１０８は音声認識の結果が発声終了のタイミングで撮像を指示する内容であるか否か判断する。
【０３３３】
発声終了のタイミングで撮像を指示する内容である場合（ステップＳ１５３６のＹＥＳ）、ステップＳ１５３７で画像Ａを消去するための信号を出力する。
【０３３４】
発声終了のタイミングで撮像を指示する内容でない場合（ステップＳ１５３６のＮＯ）、ステップＳ１５３８で、画像Ａ、画像Ｂを消去するための信号を出力する。
【０３３５】
次に、ステップＳ１５３９で、認識結果処理部１０８は音声認識の結果が発声開始のタイミングから一定時間経過後に撮像を指示する内容であるか否か判断する。
【０３３６】
発声開始から一定時間経過後に撮像を指示する内容である場合（ステップＳ１５３９のＹＥＳ）、ステップＳ１５４０で一定時間経過後（このタイミングを第３の時刻とする）に、撮像制御部１２３は撮像装置１６０３に撮像動作を実行させるための信号を出力する。
【０３３７】
尚、ステップＳ１５４０で出力された信号によって撮像された画像を画像Ｃとする。
【０３３８】
次に、ステップＳ１５４１で、画像記憶制御部１０４は、第３の保持として、直前のステップＳ１５４０で撮像された画像Ｃを表す画像データを格納装置（画像記憶用）１６１０に記憶させる信号を出力して、処理を終了する。
【０３３９】
また、発声開始のタイミングから一定時間経過後に撮像を指示する意味内容でない場合（ステップＳ１４３９のＮＯ）、処理を終了する。
【０３４０】
このような構成とすることで、発声区間に対して、第１の関係である発声開始のタイミングで撮像された第１の画像（画像Ａ）と、第２の関係である発声終了のタイミングで撮像された第２の画像（画像Ｂ）とを得ることができる。
【０３４１】
また、発声区間に対して、第３の関係である発声終了から一定時間のタイミングで撮像された第３の画像（画像Ｃ）を得ることができる。

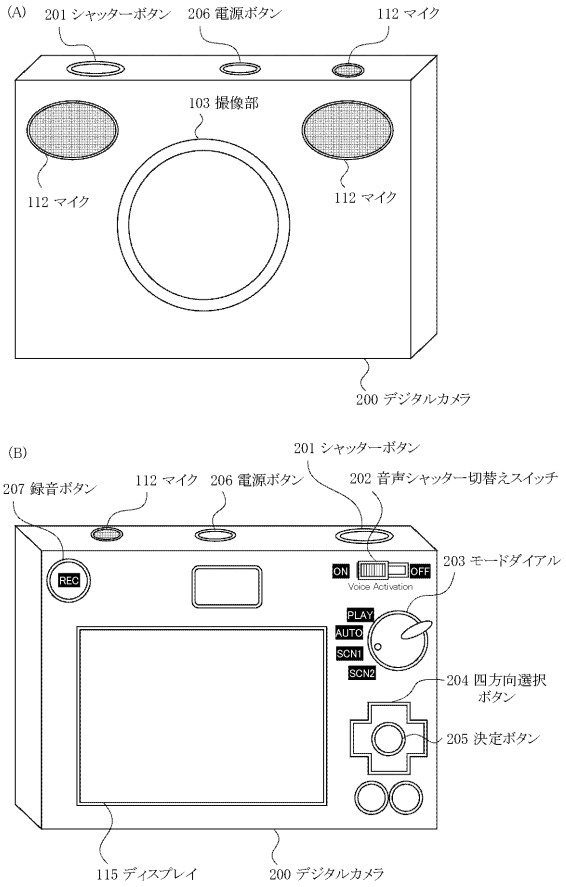
【０３４２】
さらに、音声区間の音声の意味内容に応じて、複数の画像からユーザが所望するタイミングで撮像された画像を選択することができる。
【０３４３】
また、このような構成とすることで、本実施形態の情報処理装置１６００と、外部機器とを連動させて、ユーザが欲するタイミングで撮像された画像を効率良く取得することができる。
【０３４４】
また、本実施形態の情報処理装置１６００によると、断続的に音声が入力された場合にも、１つのコマンドとして認識することが可能であるため、発声区間が長くなるような言葉をコマンドとして利用した場合にも認識の誤りが軽減される。
【０３４５】
（プログラムＣＬのサポート）
尚、本発明の目的は、前述した実施の形態の機能を実現するソフトウェアのプログラムコードを記録した記憶媒体を、システムあるいは装置に供給し、そのシステムあるいは装置のコンピュータがプログラムコードを読み出し実行することによっても達成される。
【０３４６】
尚、コンピュータは、ＣＰＵ、ＭＰＵ等であってもよい。
【０３４７】
この場合、記憶媒体から読み出されたコンピュータ読み取り可能なプログラムコード自体が前述した実施形態の機能を実現することになり、そのプログラムコードを記憶した記憶媒体は本発明を構成することになる。
【０３４８】
プログラムコードを供給するための記憶媒体としては、例えば、フレキシブルディスク，ハードディスク，光ディスク，光磁気ディスク，ＣＤ－ＲＯＭ，ＣＤ－Ｒ，磁気テープ，不揮発性のメモリカード，ＲＯＭ等を用いることができる。
【０３４９】
また、コンピュータが読出したプログラムコードを実行することにより、前述した実施形態の機能が実現されるだけでなく、そのプログラムコードの指示に基づき、ＯＳ（オペレーティングシステム）等が実際の処理の一部または全部を行ってもよい。
【０３５０】
尚、この処理によって前述した実施形態の機能が実現される場合も含まれる。
尚、ＯＳはコンピュータ上で稼働している。
【０３５１】
また、まず記憶媒体から読出されたプログラムコードが、コンピュータに挿入された機能拡張ボードやコンピュータに接続された機能拡張ユニットに備わるメモリに書込まれる。
【０３５２】
その後、そのプログラムコードの指示に基づき、その機能拡張ボードや機能拡張ユニットに備わるＣＰＵ等が実際の処理の一部または全部を行い、その処理によって前述した実施形態の機能が実現される場合も含まれる。
【図面の簡単な説明】
【０３５３】
【図１】本発明の第１の実施形態に係る情報処理装置の構成の一例を示す機能ブロック図である。
【図２】本発明の第１の実施形態で想定されるデジタルカメラの外観を示す図である。
【図３】音声検出部１０６が判定した認識状態の一例を示す図である。
【図４】音声検出部１０６の動作の一例を示す概念図である。
【図５】音声検出部１０６における処理動作を示すフローチャートである。
【図６】音声によって撮像を指示する場合のデジタルカメラ２００における処理の一例を

示す第一のフローチャートである。
【図７】音声によって撮像を指示する場合のデジタルカメラ２００における処理の一例を示す第二のフローチャートである。
【図８】音声によって撮像を指示する場合のデジタルカメラ２００における処理の一例を示す第三のフローチャートである。
【図９】本発明の第１の実施形態で利用する音声認識文法の一例を示す図である。
【図１０】認識結果制御テーブルの一例を示す図である。
【図１１】本発明の第１の実施形態に係るデジタルカメラ２００を利用して、“Ｓｈｏｏｔ”という音声指示で撮像する場合の動作を示す図である。
【図１２】本発明の第１の実施形態に係るデジタルカメラ２００を利用して、“Ｃｈｅｅｓｅ”という音声指示で撮像する場合の動作を示す図である。
【図１３】発声開始を検出した時点でのみ撮像する場合のフローチャートである。
【図１４】情報処理装置１６００における処理動作の一例を示した第１のフローチャートである。
【図１５】情報処理装置１６００における処理動作の一例を示した第２のフローチャートである。
【図１６】本発明の第２の実施形態に係る情報処理装置１６００の構成の一例を示す機能ブロック図である。
【図１７】音声検出部１０６が判定した認識状態と撮像部１０３、画像記憶制御部１０４の動作の一例を示す図である。
【符号の説明】
【０３５４】
１０１　制御部
１０４　画像記憶制御部
１０５　音声入力部
１０６　音声検出部
１０７　音声認識部
１０８　認識結果処理部
１０９　表示制御部
１２２　操作制御部
１２３　撮像制御部

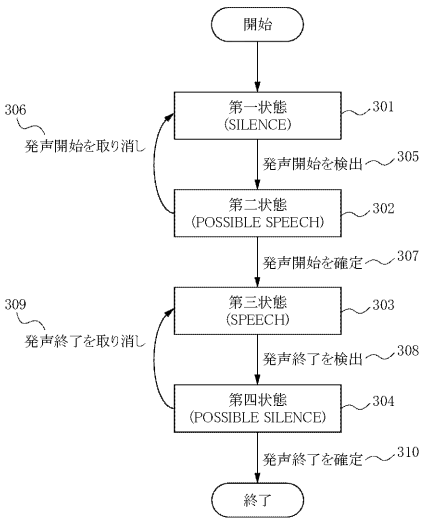
【図１】


101 制御部
200 デジタルカメラ
122
操作制御部
操作部
102
123
撮像制御部
撮像部
103
110
104
画像記憶
制御部
メモリ
（画像記憶用）
111
記憶媒体
（画像記憶用）
105
音声入力部
マイク
112
106
音声検出部
113
107
音声認識部
メモリ
（音声認識データ用）
114
108
認識結果
処理部
メモリ
（認識結果制御テーブル用）
109
表示制御部
ディスプレイ
115

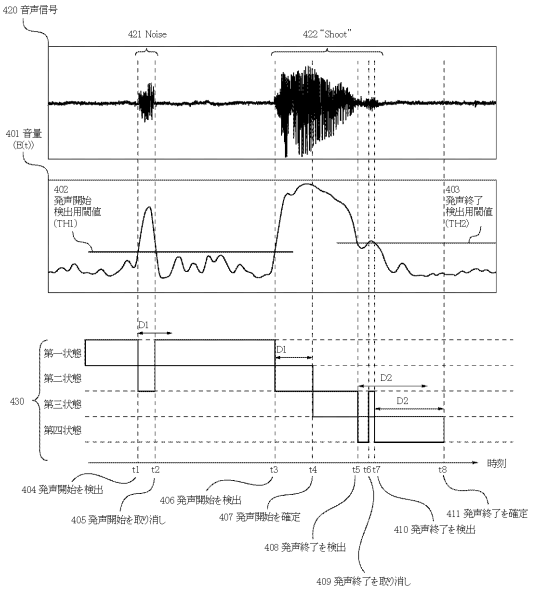
【図２】


(A)
201 シャッターボタン
206 電源ボタン
112 マイク
103 撮像部
112 マイク
112 マイク
200 デジタルカメラ
(B)
201 シャッターボタン
207 録音ボタン
112 マイク
206 電源ボタン
202 音声シャッター切替えスイッチ
REC
ON
OFF
Voice Activation
203 モードダイアル
PLAY
AUTO
SCN1
SCN2
204 四方向選択
ボタン
205 決定ボタン
115 ディスプレイ
200 デジタルカメラ

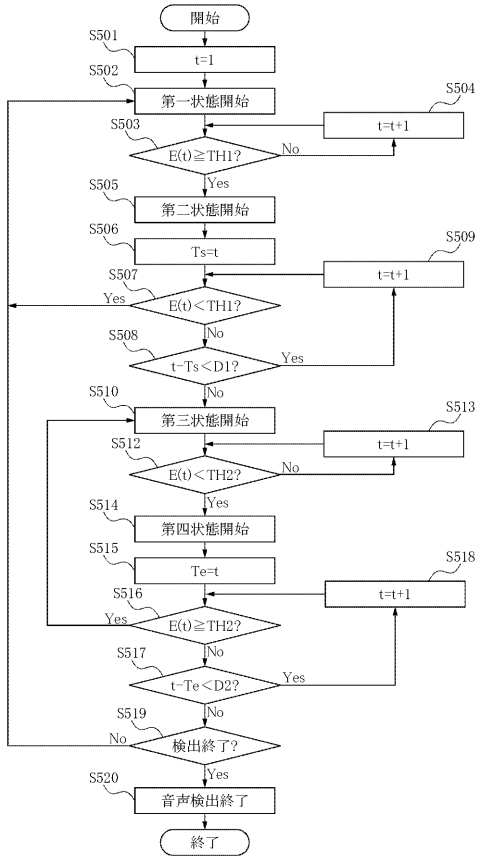
【図３】


開始
第一状態
(SILENCE)
301
306
発声開始を取り消し
発声開始を検出
305
第二状態
(POSSIBLE SPEECH)
302
発声開始を確定
307
第三状態
(SPEECH)
303
309
発声終了を取り消し
発声終了を検出
308
第四状態
(POSSIBLE SILENCE)
304
発声終了を確定
310
終了

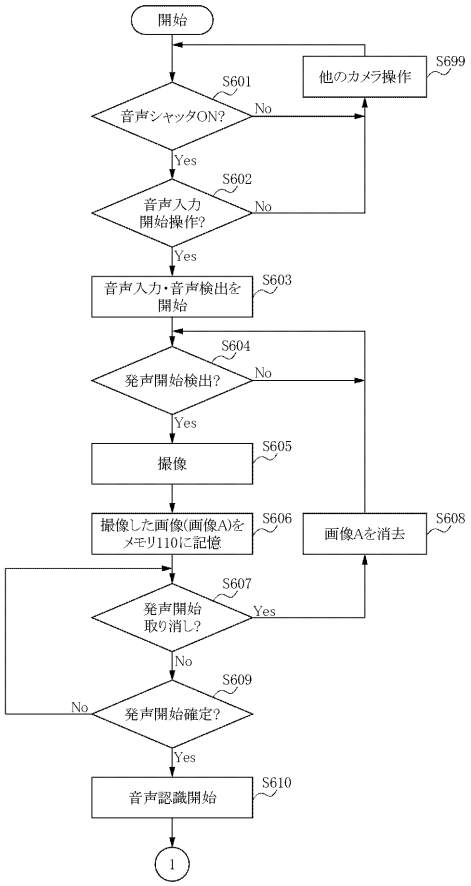
【図４】


420 音声信号
421 Noise
422 "Shoot"
401 音量
(Et)
402
発声開始
検出用閾値
(TH1)
403
発声終了
検出用閾値
(TH2)
D1
D1
D2
D2
430
第一状態
第二状態
第三状態
第四状態
時刻
t1
t2
t3
t4
t5
t6
t7
t8
404 発声開始を検出
405 発声開始を取り消し
406 発声開始を検出
407 発声開始を確定
408 発声終了を検出
409 発声終了を取り消し
410 発声終了を検出
411 発声終了を確定

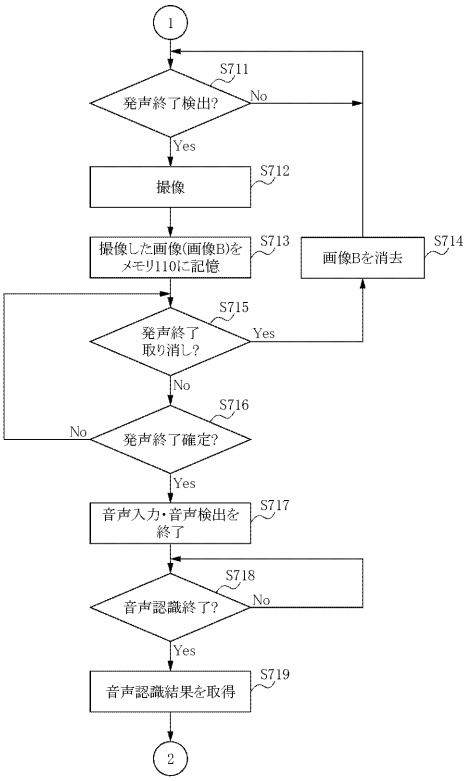
【図５】


開始
S501
t=1
S502
第一状態開始
S504
t=t+1
S503
E(t)≧TH1?
No
Yes
S505
第二状態開始
S506
Ts=t
S509
t=t+1
S507
Yes
E(t)＜TH1?
No
S508
t-Ts＜D1?
Yes
No
S510
第三状態開始
S513
t=t+1
S512
E(t)＜TH2?
No
Yes
S514
第四状態開始
S515
Te=t
S518
t=t+1
S516
Yes
E(t)≧TH2?
No
S517
t-Te＜D2?
Yes
No
S519
No
検出終了?
Yes
S520
音声検出終了
終了

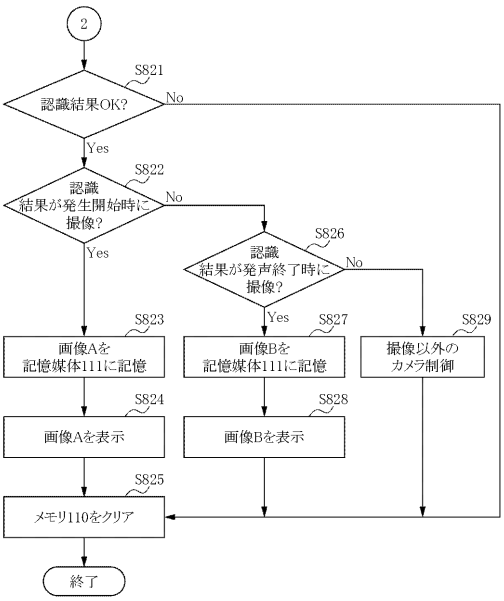
【図６】


開始
S699
他のカメラ操作
S601
音声シャッタON?
No
Yes
S602
音声入力
開始操作?
No
Yes
音声入力・音声検出を
開始
S603
S604
発声開始検出?
No
Yes
撮像
S605
撮像した画像(画像A)を
メモリ110に記憶
S606
画像Aを消去
S608
S607
発声開始
取り消し?
Yes
No
S609
No
発声開始確定?
Yes
音声認識開始
S610
1

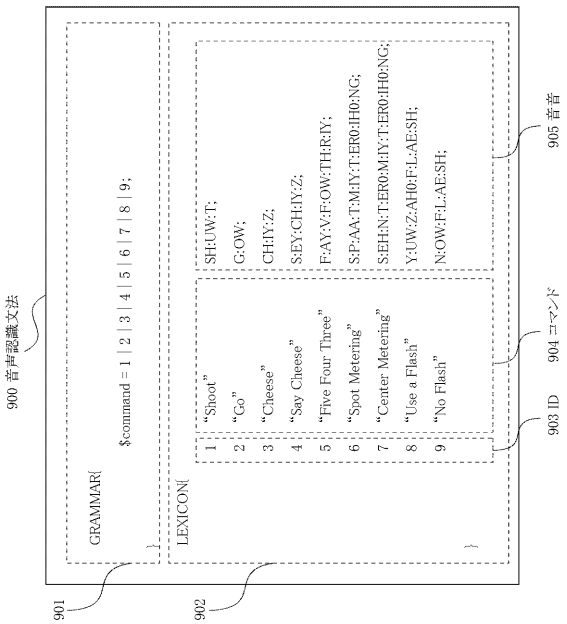
【図７】


1
S711
発声終了検出?
No
Yes
撮像
S712
撮像した画像(画像B)を
メモリ110に記憶
S713
画像Bを消去
S714
S715
発声終了
取り消し?
Yes
No
S716
No
発声終了確定?
Yes
音声入力・音声検出を
終了
S717
S718
音声認識終了?
No
Yes
音声認識結果を取得
S719
2

【図８】

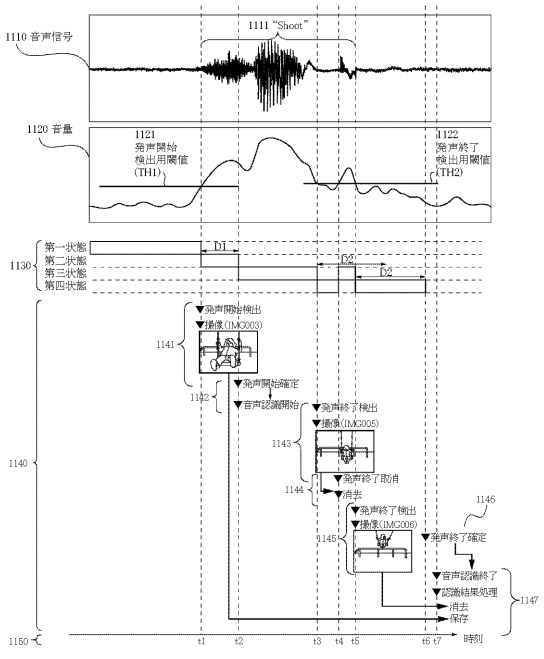

2
S821
認識結果OK?
No
Yes
S822
認識
結果が発生開始時に
撮像?
No
Yes
S826
認識
結果が発声終了時に
撮像?
No
Yes
S823
画像Aを
記憶媒体111に記憶
S827
画像Bを
記憶媒体111に記憶
S829
撮像以外の
カメラ制御
S824
画像Aを表示
S828
画像Bを表示
S825
メモリ110をクリア
終了

【図９】

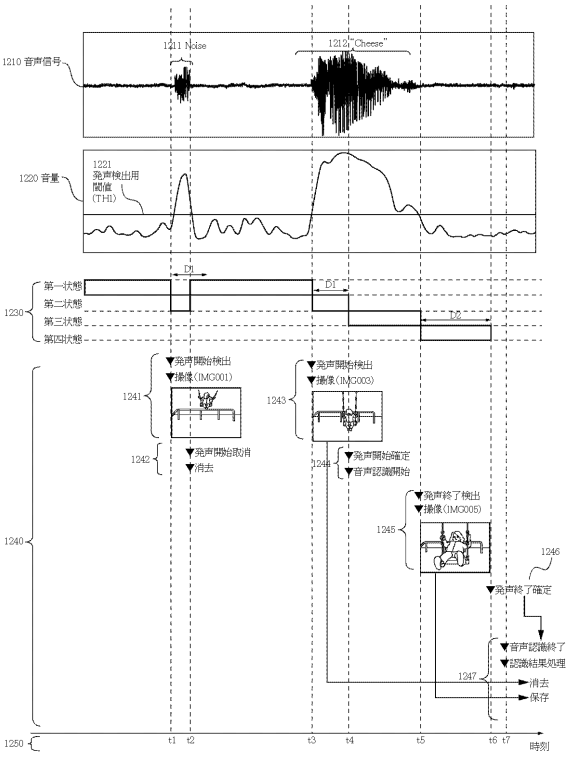


【図１０】

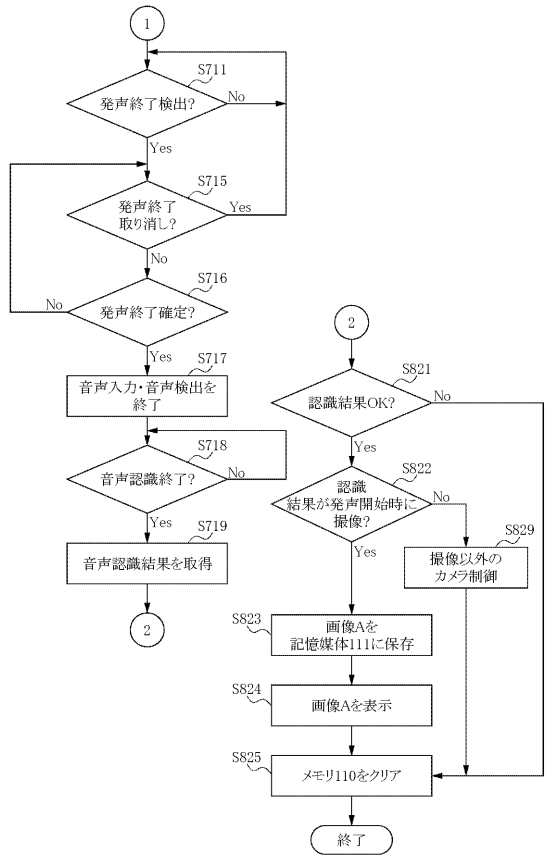
【図１１】



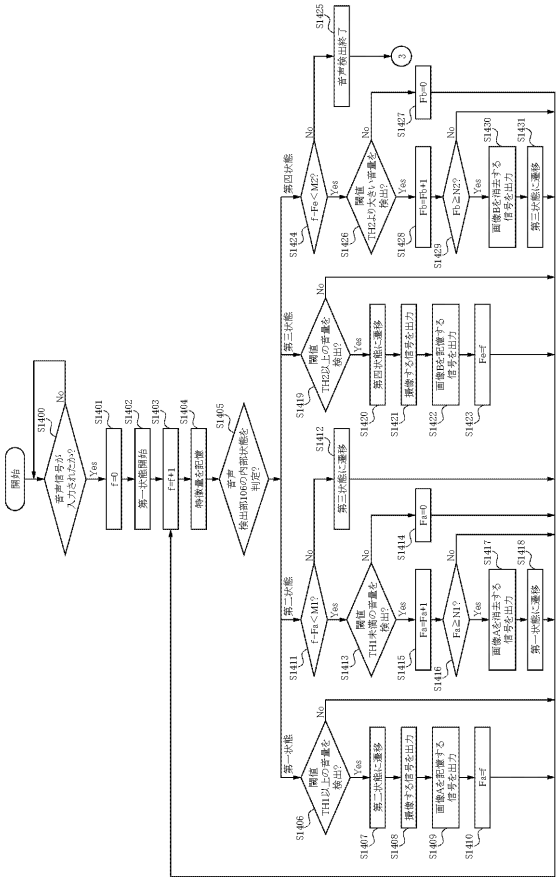
【図１２】



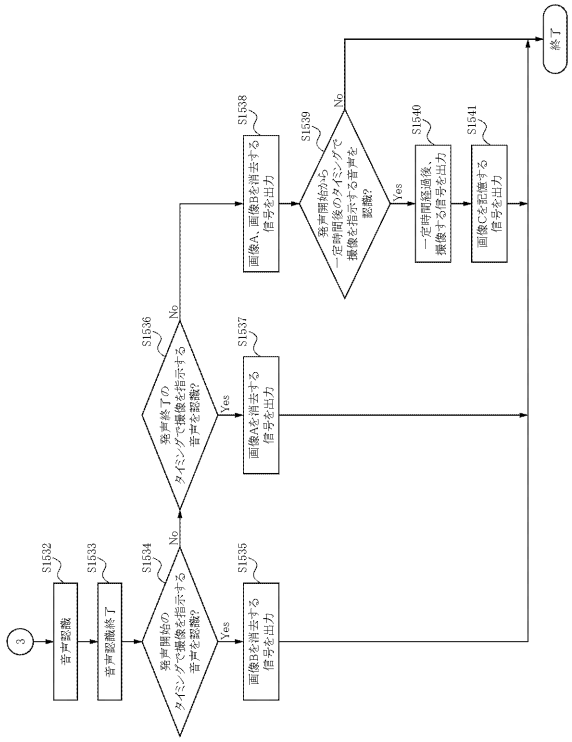
【図１３】



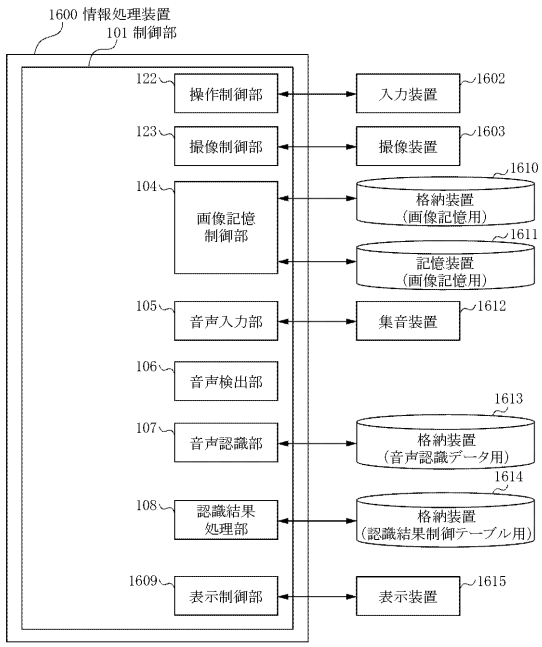
【図１４】



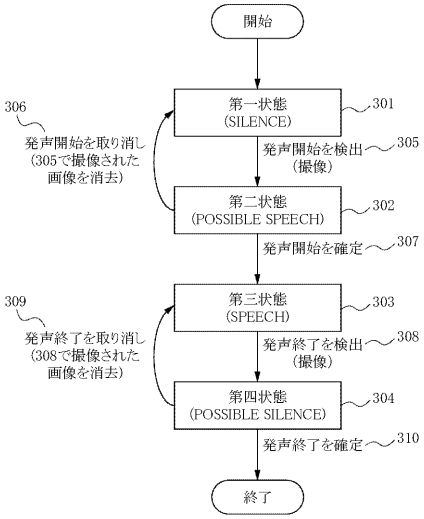
【図１５】



【図１６】



【図１７】


開始
第一状態
(SILENCE)
301
306
発声開始を取り消し
(305で撮像された
画像を消去)
発声開始を検出
(撮像)
305
第二状態
(POSSIBLE SPEECH)
302
発声開始を確定
307
第三状態
(SPEECH)
303
309
発声終了を取り消し
(308で撮像された
画像を消去)
発声終了を検出
(撮像)
308
第四状態
(POSSIBLE SILENCE)
304
発声終了を確定
310
終了

フロントページの続き

(56)参考文献 特開２００５－１８１３６５（ＪＰ，Ａ）
特開２００４－３０１８９５（ＪＰ，Ａ）
特開２００４－２８７０６３（ＪＰ，Ａ）
特開２００５－１５９５５８（ＪＰ，Ａ）
特開２００５－１２２１３９（ＪＰ，Ａ）
特開平１１－１９４３９２（ＪＰ，Ａ）
特開２００６－１８４５８９（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｈ０４Ｎ　　５／２２５