

(19) 日本国特許庁(JP)

## (12) 特許公報(B2)

(11) 特許番号

特許第5247384号  
(P5247384)

(45) 発行日 平成25年7月24日(2013.7.24)

(24) 登録日 平成25年4月19日(2013.4.19)

(51) Int.Cl.	F 1
HO4N 5/225 (2006.01)	HO4N 5/225 F
G10L 15/00 (2013.01)	G10L 15/00 200G
G10L 15/20 (2006.01)	G10L 15/20 360Z
G10L 15/28 (2013.01)	G10L 15/28 500
	G10L 15/28 230J

請求項の数 16 (全 21 頁)

(21) 出願番号 特願2008-304598 (P2008-304598)  
 (22) 出願日 平成20年11月28日 (2008.11.28)  
 (65) 公開番号 特開2010-130487 (P2010-130487A)  
 (43) 公開日 平成22年6月10日 (2010.6.10)  
 審査請求日 平成23年11月22日 (2011.11.22)

(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  
 (72) 発明者 奥谷 泰夫  
 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 藤原 敏利

最終頁に続く

(54) 【発明の名称】撮像装置、情報処理方法、プログラムおよび記憶媒体

## (57) 【特許請求の範囲】

## 【請求項 1】

ファインダとディスプレイとを備える撮像装置であって、  
 前記ファインダに画像を表示する表示モードと、前記ディスプレイに画像を表示する表示モードとを切替える切替え手段と、  
 前記ファインダに画像を表示している場合、音声入力の処理を接話用のモードに設定し、前記ディスプレイに画像を表示している場合、音声入力の処理を非接話用のモードに設定する設定手段と、

設定された音声入力のモードに従って、予め設定された制御コマンドを音声入力する入力手段とを備える撮像装置。

## 【請求項 2】

前記接話用の入力モードは、前記非接話用の入力モードよりも入力音量が小さくなるように設定することを特徴とする請求項 1 に記載の撮像装置。

## 【請求項 3】

更に、前記ファインダに画像を表示している場合、音声認識の処理を接話用のモードに決定し、前記ディスプレイに画像を表示している場合、音声認識の処理を非接話用のモードに決定する決定手段と、

決定された音声認識のモードに従って、音声入力された制御コマンドを音声認識する音声認識手段とを備える請求項 1 または請求項 2 に記載の撮像装置。

## 【請求項 4】

前記非接話用の認識モードは、前記接話用の認識モードよりも雑音が多い環境に適した設定であることを特徴とする請求項3に記載の撮像装置。

**【請求項5】**

更に、前記ファインダに画像を表示している場合に、撮像モードであるか再生モードであるか判断する判断手段と、

入力された制御コマンドに応じて、予め設定された撮像または再生に関する制御を実行する制御手段とを備え、

前記決定手段は、前記撮像モードである場合、非接話用のモードを非接話撮像用のモードに決定し、前記再生モードである場合、非接話用のモードを非接話再生用のモードに決定することを特徴とする請求項3または請求項4に記載の撮像装置。 10

**【請求項6】**

非接話撮像用のモードでは、再生に関する制御コマンドが音声認識された場合には予め設定された制御を実行せず、非接話再生用のモードでは、撮像に関する制御コマンドが音声認識された場合には予め設定された制御を実行しないことを特徴とする請求項5に記載の撮像装置。

**【請求項7】**

ファインダとディスプレイとを備える撮像装置に実行させる情報処理方法であって、

切替え手段が、前記ファインダに画像を表示する表示モードと、前記ディスプレイに画像を表示する表示モードとを切替える切替え工程と、

設定手段が、前記ファインダに画像を表示している場合、音声入力の処理を接話用のモードに設定し、前記ディスプレイに画像を表示している場合、音声入力の処理を非接話用のモードに設定する設定工程と、 20

入力手段が、設定された音声入力のモードに従って、予め設定された制御コマンドを音声入力する入力工程とを有する情報処理方法。

**【請求項8】**

請求項7に記載の情報処理方法をコンピュータに実行されるためのプログラム。

**【請求項9】**

請求項8に記載のプログラムを記憶した記憶媒体。

**【請求項10】**

ファインダとディスプレイとを備える撮像装置であって、

前記ディスプレイに画像を表示しないモードと、前記ディスプレイに画像を表示するモードを切替える切替え手段と、

前記ディスプレイに画像が表示されていない場合、音声入力の処理を接話用のモードに設定し、前記ディスプレイに画像が表示されている場合、音声入力の処理を非接話用のモードに設定する設定手段と、

設定された音声入力のモードに従って、予め設定された制御コマンドを音声入力する入力手段とを備える撮像装置。 30

**【請求項11】**

前記接話用の入力モードは、前記非接話用の入力モードよりも入力音量が小さくなるように設定することを特徴とする請求項10に記載の撮像装置。 40

**【請求項12】**

更に、前記ディスプレイに画像を表示していない場合、音声認識の処理を接話用のモードに決定し、前記ディスプレイに画像を表示している場合、音声認識の処理を非接話用のモードに決定する決定手段と、

決定された音声認識のモードに従って、音声入力された制御コマンドを音声認識する音声認識手段とを備える請求項10または請求項11に記載の撮像装置。

**【請求項13】**

前記非接話用の認識モードは、前記接話用の認識モードよりも雑音が多い環境に適した設定であることを特徴とする請求項12に記載の撮像装置。

**【請求項14】**

ファインダとディスプレイを備える撮像装置に実行させる情報処理方法であって、切替え手段が、前記ディスプレイに画像を表示しないモードと、前記ディスプレイに画像を表示するモードを切替える切替え工程と、

設定手段が、前記ディスプレイに画像が表示されていない場合、音声入力の処理を接話用のモードに設定し、前記ディスプレイに画像が表示されている場合、音声入力の処理を非接話用のモードに設定する設定工程と、

入力手段が、設定された音声入力のモードに従って、予め設定された制御コマンドを音声入力する入力工程とを有する情報処理方法。

【請求項 1 5】

請求項 1 4 に記載の情報処理方法をコンピュータに実行されるためのプログラム。

10

【請求項 1 6】

請求項 1 5 に記載のプログラムを記憶した記憶媒体。

【発明の詳細な説明】

【技術分野】

【0 0 0 1】

本発明は、音源の位置を推定して、音声入力に関する設定を切替える技術に関する。

【背景技術】

【0 0 0 2】

従来、右目又は左目の一方で覗くようにして使用するファインダと、目から所定の距離だけ離間させ、かつ正対させて使用するディスプレイとを備えるデジタルカメラが知られている。

20

【0 0 0 3】

従って、上記のデジタルカメラを利用して被写体を撮像する場合、ユーザはファインダを覗きながら、撮像するための操作をすることもでき、ディスプレイを見ながら、撮像するための操作をすることもできる。

【0 0 0 4】

一方、マイクロフォンを備えるデジタルカメラが存在する。

そして、上記のマイクロフォンに入力された音を、撮像された画像に対応付けて記録する技術が知られている。

【0 0 0 5】

30

また、上記マイクロフォンに入力された音をトリガーとして、デジタルカメラを操作する技術が知られている。

以下、マイクロフォンを備えるデジタルカメラを用いて、音声を入力しながら撮像する場面について説明する。

【0 0 0 6】

この場面において、ユーザはファインダを覗きながら音声入力をを行う場合と、ディスプレイを見ながら音声入力をを行う場合とが考えられる。

つまり、音源となるユーザの口とマイクロフォンとの距離は一定ではない。

従って、マイクロフォンに入力される音の音量等は不安定となる。

【0 0 0 7】

40

このような従来技術に鑑みて、マイクロフォンの感度を切替えるスイッチを備え、音源の位置にあわせて、ユーザがマイクロフォンの感度を切替えることができる技術が知られている（例えば、特許文献 1）。

【特許文献 1】特開昭 58 - 137828 号公報

【発明の開示】

【発明が解決しようとする課題】

【0 0 0 8】

しかしながら、手動でマイクロフォンの感度を切替えるのは非効率である。

また、特許文献 1 では、ユーザがマイクロフォンの感度を切替えずに、ある場面では、ファインダを覗きながら音声入力をを行い、別の場面では、ディスプレイを見ながら音声入

50

力を行うことが考える。

**【0009】**

このとき、ファインダを覗きながら音声入力を行った場合と、ディスプレイを見ながら音声入力を行った場合のどちらか一方の場合しか好適な音声入力がなされない。

**【0010】**

好適な音声入力がなされない（例えば、入力される音の音量が不安定となる）と、入力された音声を録音し、当該音声を再生する場合には、出力音量を調整する必要がある。

**【0011】**

また、入力された音声に対して音声認識の処理を行う場合には、認識精度を低下させる一因となる。

10

**【0012】**

また、音声入力をしながら撮像する際に、マイクロフォンの感度を切替えるスイッチを操作すると、一瞬のシャッターチャンスを逃す場合がある。

**【0013】**

本発明は、上述した課題に鑑みてなされたものであり、ユーザがデジタルカメラを操作する際の口（音源）の位置を推定し、推定された音源から発せられる音を好適に取得する技術を提供することを目的とする。

**【0014】**

また、録音または音声認識の処理を実行する際に、安定した音量の音（例えば、十分な音量の音、適切な音量の音等）を入力させることを目的とする。

20

**【課題を解決するための手段】**

**【0015】**

上記の課題を解決するために、本発明に係る撮像装置は、ファインダとディスプレイとを備え、前記ファインダに画像を表示する表示モードと、前記ディスプレイに画像を表示する表示モードとを切替える切替え手段と、前記ファインダに画像を表示している場合、音声入力の処理を接話用のモードに設定し、前記ディスプレイに画像を表示している場合、音声入力の処理を非接話用のモードに設定する設定手段と、設定された音声入力のモードに従って、予め設定された制御コマンドを音声入力する入力手段とを備える。

**【発明の効果】**

**【0016】**

30

本発明によれば、ユーザがデジタルカメラを操作する際の口（音源）の位置を推定し、推定された音源から発せられる音を好適に取得することが可能となる。

**【0017】**

また、録音または音声認識の処理を実行する際に、安定した音量の音を入力させることが可能となる。

**【発明を実施するための最良の形態】**

**【0018】**

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。

**【0019】**

（第1の実施形態）

40

図1は第1の実施形態に係る撮像装置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。

**【0020】**

図1において、デジタルカメラ100は、制御部101、撮像部102、記録再生部103、操作部104、表示部105、音声入力部106、音声処理部107を備える。

**【0021】**

（各部の説明）

制御部101は、撮像部102、記録再生部103、操作部104、表示部105、音声入力部106、音声処理部107の動作を制御する。

尚、制御部101における制御は後述する。

50

**【0022】**

尚、制御部101は、CPU(中央演算装置)、ROM(Read Only Memory)、RAM(Random Access Memory)等で構成される。

**【0023】**

撮像部102は、レンズによって結像した像の画像信号を生成し、生成した画像信号にアナログ/デジタル(A/D)変換等の処理を施し、記録媒体にデジタルの画像信号(以下、画像データとする)を出力する。

尚、撮像部102は、レンズ、CCDやCMOS等の撮像素子、A/D変換回路等で構成される。

記録再生部103は、記録媒体に画像データを記録させる。

10

また、記録再生部103は、記録媒体に記録された画像データを読み出す。

**【0024】**

尚、記録再生部103は、ハードディスクドライブ(HDD)やフラッシュメモリ等の記録媒体、記憶媒体に画像データを記録する記録制御部、記録媒体から画像データを読み出す再生制御部等で構成される。

**【0025】**

操作部104はユーザ(使用者)がデジタルカメラ100を手動で操作する際に使用する。

尚、操作部104は、ボタン、スイッチ等で構成される。

**【0026】**

以下、操作部104を構成するボタンの例を説明する。

20

操作部104は、音声処理操作ボタン141、表示切替ボタン142を有する。

音量処理操作ボタン141が操作された場合、音声入力部106は音の取得を開始し、音声処理部107は録音、音声認識等の処理を実行する。

表示切替ボタン142が操作された場合、画像を表示する表示モードが切替わる。

尚、表示モードの詳細は後述する。

**【0027】**

表示部105は、ファインダ150、LCD、TFT、有機ELディスプレイ等の表示ディスプレイ151等で構成され、ファインダ150と表示ディスプレイ151は以下の3種類の表示モードをとる。

30

**【0028】**

尚、ファインダ150は、周知の電子ビューファインダ(EVF)であり、ユーザが接眼して(片目で覗くようにして)画像や撮像条件等を確認する場合に好適である。

**【0029】**

また、表示ディスプレイ151は、ユーザが所定の距離(例えば、15cmから50cm程度)離間させて、画像や撮像条件等を確認する場合に好適である。

**【0030】**

次に、上述した表示モードについて説明する。

第1の表示モードとは、ファインダ150に画像や撮像条件等を表示し(ON)、表示ディスプレイ151には何も表示しない(OFF)モードである。

40

**【0031】**

第2の表示モードとは、ファインダ150に何も表示せず(OFF)、表示ディスプレイ151に画像や撮像条件等を表示する(ON)モードである。

**【0032】**

第3の表示モードとは、ファインダ150、表示ディスプレイ151の双方に画像や撮像条件等を表示する(ON)モードである。

**【0033】**

尚、上述の3つの表示モードは、例えば、ユーザが表示切替ボタン142を操作する度に、第1の表示モード、第2の表示モード、第3の表示モード、第1の表示モード、...というように順に切替わる。

50

**【0034】**

音声入力部106は、マイクロフォン160等で構成され、音を取得する処理を実行する。

**【0035】**

音声処理部107は、音声入力部106が取得した音に対して、録音、音声認識等の処理を行う。

音声処理部107が録音を行う場合、取得した音は記録媒体に記録される。

音声処理部107が音声認識を行う場合は、認識結果に基づいて、その後の処理が実行される。

**【0036】**

また、録音と音声認識のいずれの処理を実行するかは、例えば、以下のように決定する。

**【0037】**

即ち、音声処理操作ボタン141を長押し（例えば、押下した時間が1秒間以上）した場合には録音を、音声処理操作ボタン141を短押し（例えば、押下した時間が1秒間未満）した場合には音声認識を実行する。

**【0038】**

尚、本実施形態に係るデジタルカメラは1つの音声処理操作ボタン141備えるが、当該音声処理操作ボタン141に替えて、録音を実行するための録音ボタンと、音声認識を実行するための音声認識ボタンとを備えてもよい。

10

**【0039】**

以下、表示部106の表示モードに基づいて、音声入力部106、音声処理部107等の設定を切替える処理について説明する。

**【0040】**

（音声入力部106の設定を切替える処理を示すフローチャートの説明）

図2は、本実施形態に係るデジタルカメラにおいて、表示モードに基づいて音声入力の処理を切替える情報処理の流れを説明するフローチャートである。

20

**【0041】**

ステップS201において、制御部101は、音声処理操作ボタン141が押下されたか否か判断し、音声処理操作ボタン141が押下された場合、ステップS202の処理を実行する。

30

**【0042】**

一方、音声処理操作ボタン141が押下されなかった場合、音声処理操作ボタン141が押下されるまで待機する。

**【0043】**

ステップS202において、制御部101は、表示ディスプレイ151に画像や撮像条件が表示されているか否か判断し、表示ディスプレイ151に画像や撮像条件が表示されている場合、ステップS203の処理を実行する。

**【0044】**

即ち、表示部105の表示モードが第2の表示モードまたは第3の表示モードである場合、ステップS203の処理を実行する。

40

**【0045】**

ステップS203において、制御部101は、音声入力部106の動作を非接話用に設定する。

**【0046】**

何故なら、表示ディスプレイ151に画像が表示されている場合は、ユーザはカメラの表示ディスプレイ151を見ながら操作できる姿勢になると推定されるからである。

**【0047】**

つまり、図3(a)のように、ユーザの顔は表示ディスプレイ151から少し離れた位置（例えば、表示ディスプレイ151とユーザの顔（口）との間隔が15cmから50c

50

mとなるような位置)にあると推定される。

**【0048】**

そこで、音声入力部106のから少し離れた位置(例えば、表示ディスプレイ151とユーザの顔(口)との間隔が15cmから50cmとなるような位置)にある音源から発せられる音を好適に取得するような設定をする。

**【0049】**

尚、音声入力部106における非接話用(入力モード)の設定については後述する。

**【0050】**

一方、表示ディスプレイ151に画像や撮像条件が表示されていない場合、ステップS204の処理を実行する。 10

**【0051】**

即ち、表示部105の表示モードが第1の表示モードである場合、ステップS204の処理を実行する。

**【0052】**

ステップS204において、制御部101は、音声入力部106の動作を接話用に設定する。

**【0053】**

何故なら、ファインダ150に画像が表示されている場合は、ユーザはカメラのファインダ150を見ながら操作する姿勢になると推定されるからである。

**【0054】**

つまり、図3(b)のように、ユーザの顔はファインダ150と近接した位置(例えば、ファインダ150とユーザの顔(口)との間隔が5cm以下となるような位置)にあると推定される。 20

**【0055】**

そこで、音声入力部106と近接した位置(例えば、ファインダ150とユーザの顔(口)との間隔が5cm以下となるような位置)の音源から発せられる音を好適に取得するような設定をする。

尚、音声入力部106における接話用(入力モード)の設定については後述する。

**【0056】**

次にステップS205において、音声入力部106は音を取得する処理を開始する。

以下に、音声入力部106における接話用の設定と非接話用の設定について設定項目毎に説明する。 30

まず、マイクロフォン160の入力レベルを調整する場合について説明する。

入力レベルとは、例えば、音量、音圧等である。

**【0057】**

本実施形態においては、接話用の音量として音量V1、非接話用の音量として音量V2を用いる。

ここで、非接話用の音量V2は、接話用の音量V1よりも、音を大きく増幅するように設定されている。

**【0058】**

例えば、音量を0~15の16段階で調整可能で、数字が大きくなるほど入力音量が大きくなるような音声入力装置を用いる場合、音量V1を5に、音量V2を12に設定する。 40

**【0059】**

このようにすると、ユーザが手動で入力音量を切替えることなく、ユーザがカメラを操作する姿勢を反映した入力音量の切替えが実現される。

**【0060】**

すなわち、表示ディスプレイ151がONの時は、図3(a)のように、ユーザはカメラを少し離して構える姿勢になると推定して、非接話用の音量に設定することにより、離れたところから発せられた音声が安定した音量で入力される。 50

**【0061】**

また、表示ディスプレイ151がOFFの時は、ユーザはファインダ150を覗くためにカメラを顔の近くに構える姿勢になると推定して、入力音量を接話用に設定することにより、マイクロフォンの近くで発せられた音声が安定した音量で入力される。

したがって、ユーザが手動操作をすることなく、安定した音量で集音することが可能となる。

**【0062】**

次に、マイクロフォン160の指向性を設定する場合について説明する。

ユーザは、図3(a)に示すように、表示ディスプレイ151を見る姿勢をとる場合と、図3(b)に示すように、ファインダ150を覗く姿勢をとる場合がある。

10

**【0063】**

そして、それぞれの場合において、ユーザの口元であると推定される位置は異なる。

**【0064】**

従って、ユーザが表示ディスプレイ151を見ている場合、図4(a)に示すように、マイクロフォンの指向性が狭指向性であるとユーザの発した音声を好適に取得することができる。

**【0065】**

また、マイクロフォンの指向性を狭指向性とすると、周囲の雑音などユーザが発した音声以外の音を集音しにくくなるという利点もある。

**【0066】**

一方、ユーザがファインダ150を覗いている場合、図4(b)に示すように、マイクロフォンの指向性は広指向性であるとユーザの発した音声を好適に取得することができる。

20

**【0067】**

このようにすると、ユーザが手動でマイクロフォンの指向性を切替えることなく、ユーザがカメラを操作する姿勢を反映してマイクロフォンの指向性の切替えることが可能となる。

即ち、ユーザが手動操作をすることなく、ノイズ等の少ない音を取得することが可能となる。

**【0068】**

また、接話用の指向性の設定は超指向性、鋭指向性とし、非接話用の指向性の設定は単一指向性、無指向性としても同等の効果を得ることができる。

30

**【0069】**

(音声処理部107の設定を切替える処理を示すフローチャートの説明)

図5は、本実施形態に係るデジタルカメラにおいて、表示モードに基づいて、音声処理に関する種々のパラメータを設定する処理を説明するフローチャートである。

**【0070】**

ステップS501において、制御部101は、音声処理操作ボタン141が押下されたか否か判断し、音声処理操作ボタン141が押下された場合、ステップS502の処理を実行する。

40

**【0071】**

一方、音声処理操作ボタン141が押下されなかった場合、音声処理操作ボタン141が押下されるまで待機する。

**【0072】**

ステップS502において、制御部101は、表示ディスプレイ151に画像や撮像条件が表示されているか否か判断し、表示ディスプレイ151に画像や撮像条件が表示されている場合、ステップS503の処理を実行する。

**【0073】**

ステップS503において、制御部101は、音声処理部107の動作を非接話用に設定する。

50

**【0074】**

即ち、表示部105の表示モードが第2の表示モードまたは第3の表示モードである場合、ステップS503の処理を実行する。

**【0075】**

一方、表示ディスプレイ151に画像や撮像条件が表示されていない場合、ステップS504の処理を実行する。

ステップS504において、制御部101は、音声処理部107の動作を接話用に設定する。

**【0076】**

即ち、表示部105の表示モードが第1の表示モードである場合、ステップS504の処理を実行する。 10

音声処理部107における接話用の設定と非接話用の設定については後述する。

次にステップS505において、音声処理部107は音声入力部106によって取得された音に対して予め設定された音声処理を実行する。

**【0077】**

以下に、音声処理部107における接話用の設定と非接話用の設定について、音声処理部107が音声認識処理を行う場合について設定項目毎に説明する。

まず、音声認識の動作パラメータを調整する場合について説明する。

音声認識の動作パラメータとは、例えば、音声区間検出の閾値である。

このとき、接話用の設定として接話用閾値TH1、非接話用の設定として非接話用閾値TH2を用いる。 20

**【0078】**

本実施形態においては、非接話用閾値TH2は接話用閾値TH1よりも小さな音声を検出できる値とする。

例えば、音量を基準に音声区間検出を行う場合、TH1を60dB、TH2を50dBとする。

**【0079】**

このようにすると、ユーザが手動で音声区間検出の閾値を切替えることなく、ユーザがカメラを操作する姿勢を反映した音声区間検出の閾値が設定される。

また、音声区間の検出の閾値を切替えると、一般に音声認識の認識性能が向上するという効果が得られる。 30

**【0080】**

次に、接話用と非接話用とで使用する音響モデルを替える場合について説明する。

このとき、接話用の音響モデルとして音響モデルAM1を用い、非接話用の音響モデルとして音響モデルAM2を用いる。

**【0081】**

尚、音響モデルAM1はS/N比（信号対雑音比）が大きい（例えば、15～30dBとする）学習データで学習させたモデルであり、音響モデルAM2はS/N比が小さい（例えば、0～15dBとする）学習データで学習されたモデルである。

このようにすると、一般に音声認識の認識性能が向上するという効果が得られる。 40

**【0082】**

次に、音声認識の探索条件を変更する場合について説明する。

音声認識の処理においては、探索処理を高速化するため、処理の過程で不要と判断した仮説を間引く周知の枝刈り（プルーニング）の処理が行われる。

**【0083】**

代表的な処理手法は、最もスコアの大きい仮説のスコアから一定値を引いた値を閾値として、閾値未満の仮説については以降の演算を行わない方法である。

つまり、最大スコアからの差が閾値以内のスコアになる仮説のみ演算を行う。

**【0084】**

しかしながら、このような枝刈りの閾値は、音声入力時の雑音環境によって最適値が変 50

わることがよく知られている。

入力された音声に含まれる雑音が多いと探索の過程で本来正解である仮説のスコアが小さくなる。

#### 【0085】

このため、枝刈りの条件が厳しい（上記の例では閾値が小さい）と探索の過程で本来正解の仮説まで間引いてしまい、認識性能が低下する。

一方、雑音が多い環境に合わせて枝刈りの条件を緩く（上記の例では閾値を大きく）すると、雑音が少ない環境では不要な仮説の演算が発生する。

したがって、周囲の雑音を反映して枝刈りの閾値を適切に設定することが望ましい。

#### 【0086】

以上踏まえ、本実施形態においては、マイクロフォン近くで発声するため S / N 比が高くなる接話用の探索条件として条件 C1 を用い、S / N 比が低くなる非接話用の探索条件として条件 C1 よりも緩い条件 C2 を用いる。

#### 【0087】

例えば、音声認識の探索処理に、探索の過程でスコアの高い方から上位 N 個以外の仮説を枝刈りするヒストグラム・ブルーニングを適用する場合に、条件 C1 には  $N = 500$  、条件 C2 には  $N = 1000$  を用いる。

#### 【0088】

枝刈りによって残す仮説数が多い条件 C2 の方が条件 C1 よりも探索条件が緩くなっている。

このようにすると、音声認識の認識性能が向上するという効果が得られる。

#### 【0089】

（第1の実施形態の変形例）

図6は、本実施形態に係るデジタルカメラにおける音声入力部106の動作を切替える処理の変形例を説明するフローチャートである。

#### 【0090】

ステップ S601において、制御部101は、表示ディスプレイ151に画像や撮像条件が表示されているか否か判断し、表示ディスプレイ151に画像や撮像条件が表示されている場合、ステップ S602の処理を実行する。

#### 【0091】

ステップ S602において、制御部101は、音声処理部106の動作を非接話用に設定する。

#### 【0092】

一方、表示ディスプレイ151に画像や撮像条件が表示されていない場合、ステップ S603の処理を実行する。

#### 【0093】

ステップ S603において、制御部101は、音声処理部106の動作を接話用に設定する。

#### 【0094】

ステップ S604において、制御部101は、音声処理操作ボタン141が押下されたか否か判断し、音声処理操作ボタン141が押下された場合、ステップ S605の処理を実行する。

#### 【0095】

一方、音声処理操作ボタン141が押下されなかった場合、制御部101は、音声処理操作ボタン141が押下されるまで待機する。

#### 【0096】

制御部101がこのような処理を実行すると、表示部150の表示モードが変化した時に、音声処理に先立って音声処理部106の動作を設定するため、音声処理ボタン141が押下と同時に、ユーザがカメラを操作する姿勢を反映した適切な音声処理が開始される。

10

20

30

40

50

**【0097】**

尚、図6の処理は、図5のステップS501と、ステップS502からステップS504までの処理の順序を入れ替えたものである。

**【0098】**

図2に示した音声処理部107における処理も、同様に、処理の順序を入れ替えてもいい。

**【0099】**

## (第2の実施形態)

本実施形態においては、操作部104として、更に、撮像モードと再生モードとを切替えるモード切替ボタン143を備えるデジタルカメラについて説明する。

10

**【0100】**

本実施形態に係るデジタルカメラにおいてモード切替ボタン143が操作された場合、周知の撮像モード(画像を撮像するモード)、再生モード(カメラ内に記憶している画像を再生するモード)等に切替わる。

**【0101】**

本実施形態においては、撮像モードである場合、音声入力部106、音声処理部107を接話用に設定し、再生モードである場合、音声入力部106、音声処理部107を非接話用に設定する。

**【0102】**

即ち、図5のステップS501の処理、図6のステップS601の処理に替えてデジタルカメラのモードが撮像モードであるか判断する処理を実行することに相当する。

20

**【0103】**

例えば、デジタル一眼レフカメラの一部の機種では、ファインダを覗いて、被写体を確認しながら撮像する場合がある。

**【0104】**

一方、当該デジタル一眼レフカメラを用いて画像の再生を行う場合、ユーザは表示ディスプレイを見ながら画像を確認する。

**【0105】**

即ち、撮像モードにおいてユーザが音声入力をする場合、音源はデジタル一眼レフカメラ本体と近接している(数センチメートル程度とする)。

30

**【0106】**

一方、再生モードにおいてユーザが音声入力をする場合、その音源は、デジタル一眼レフカメラ本体に対して一定の間隔(数十センチメートル程度とする)がある。

**【0107】**

従って、ユーザの撮像姿勢から推定される音源の位置に対して好適な音声入力、音声処理が可能となる。

**【0108】**

## (第3の実施形態)

本実施形態に係るデジタルカメラにおいては、モード切替ボタン143が操作された場合、撮像モードか再生モードのいずれかに切替わる(第1の切替え)。

40

**【0109】**

更に、撮像モードにおいて、表示切替ボタン142が操作された場合、更に表示ディスプレイのON/OFFが切替わる(第2の切替え)。

**【0110】**

本実施形態に係るデジタルカメラは、ユーザがデジタルカメラを操作する際の口(音源)の位置を推定し、推定された音源から発せられる音を好適に取得する技術を提供することを目的とする。

**【0111】**

また、音声認識の処理を実行する際に、安定した音量の音を入力させることを目的とする。

50

**【0112】**

更に、撮像、再生等の状況に応じて音声入力される制御コマンドを効率良く判定することを目的とする。

**【0113】**

本実施形態に係るデジタルカメラは、ファインダを覗きながら撮像する場合に入力された音声を好適に入力し、認識する第1のモードと、表示ディスプレイを見ながら撮像する場合に入力された音声を好適に入力し、認識する第2のモードとを有する。

**【0114】**

また、表示ディスプレイに画像を表示して再生する場合に入力された音声を好適に入力し、認識する第3のモードを有する。 10

**【0115】**

図7は、本実施形態に係るデジタルカメラにおいて、各々の入力及び認識モードを切替える処理を説明するフローチャートである。

**【0116】**

ステップS701において、制御部101は、撮像モードであるか否か判断し、撮像モードである場合、ステップS702の処理を実行する。

**【0117】**

一方、撮像モードでない（再生モードである）場合、ステップS703においては、入力及び認識モードを非接話再生用とする。非接話再生用の認識モードの詳細は後述する。 20

**【0118】**

また、ステップS702において、制御部101は、表示ディスプレイ151がOFFであるか否か判断し、OFFである場合、ステップS704の処理を実行し、ONである場合、ステップS705の処理を実行する。

**【0119】**

ステップS704においては、入力及び認識モードを接話撮像用とする。接話撮像用の入力及び認識モードの詳細は後述する。

**【0120】**

また、ステップS705においては、入力及び認識モードを非接話撮像用とする。非接話撮像用の入力及び認識モードの詳細は後述する。 30

**【0121】**

次にステップS706において、音声入力部106は音を取得する処理を開始する。

**【0122】**

次にステップS707において、音声処理部107は音声入力部106によって取得された音に対して予め設定された音声認識の処理を実行する。

**【0123】**

以下、接話撮像用、非接話撮像用、非接話再生用の入力及び認識モードについて説明する。

**【0124】**

3つの入力及び認識モードは、接話用と非接話用に分類することができる。 40

**【0125】**

そして、この分類に基づいて、第1の実施形態と同様に、音量、指向性、音声区間検出の閾値、

音響モデル、探索条件等を設定する。

**【0126】**

更に、本実施形態に係るデジタルカメラにおいては、3つの入力及び認識モードのそれについて、音声入力された制御コマンドに応じて実行する処理を変更することを特徴とする。

**【0127】**

図8は、音声入力された制御コマンドに応じて実行する処理を示すデータテーブルを表す図である。 50

## 【0128】

本実施形態においては、接話撮像用の場合、制御コマンドを音声入力することによって撮像や機器の設定に関する多様な制御を実行できるようとする。

## 【0129】

また、非接話撮像用の場合、制御コマンドを音声入力することによって撮像や機器の設定に関する限られた制御を実行できるようとする。

## 【0130】

これは、音声入力ではなく、手動による操作入力によっても撮像や機器の設定に関する制御を実行できるからである。

## 【0131】

非接話撮像用の場合、周囲雑音の影響を受け易いことから、認識可能な制御コマンドを制限すると音声認識の精度が向上する。

10

## 【0132】

また、非接話再生用の場合、制御コマンドを音声入力することによって再生や機器の設定に関する限られた制御を実行できるようとする。

## 【0133】

このようにすると、非接話再生用の場合も非接話撮像用の場合と同様に、音声認識の精度が向上する。

## 【0134】

また、再生モードにおいては、デジタルカメラから音声出力がなされる場合がある。

20

## 【0135】

従って、非接話再生用の場合は、非接話撮像用の場合よりも、制御コマンドを特異的に音声認識可能な設定にしておくとよい。

## 【0136】

(第4の実施形態)

第1の実施形態、第2の実施形態においては、操作部104が操作された場合に、音声入力部106、音声処理部107の設定を切替えるデジタルカメラについて説明した。

## 【0137】

一方、本実施形態においては、ファインダ150に、ユーザの視線を検出する周知のセンサを備える。

30

## 【0138】

本実施形態に係るデジタルカメラにおいて、センサが視線を検知した場合、音声入力部106、音声処理部107を接話用に設定し、視線を検知しなかった場合、音声入力部106、音声処理部107を非接話用に設定する。

## 【0139】

即ち、図2のステップS202の処理、図5のステップS502の処理、図6のステップ601の処理に替えてセンサが視線を検知したか否か判断する処理を実行することに相当する。

## 【0140】

このようにすることによって、ユーザの口(音源)とマイクロフォン160との位置関係に基づいて、音声入力部106、音声処理部107の設定を切替えることが可能となる。

40

## 【0141】

図9は第4の実施形態に係る撮像装置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。図1と共に通の要素には同一の符号を付し、その説明を省略する。

## 【0142】

図9において、デジタルカメラ900は、制御部101、撮像部102、記録再生部103、操作部104、表示部105、音声入力部706、音声処理部107を備える。

## 【0143】

(各部の説明)

音声入力部906は、接話用マイクロフォン961、非接話用マイクロフォン962等

50

で構成される。

【0144】

接話用マイクロフォン961は、当該マイクロフォンからの距離が近い（例えば、10cm以内）音源から発生する音を取得する場合に好適なマイクロフォンである。

【0145】

これは、ユーザがファインダ150を覗きながら発声する場合等、ユーザの口（音源）とデジタルカメラ900との距離が数センチメートルになるからである。

【0146】

また、接話用マイクロフォン961はユーザがファインダ150を覗いたとき、ユーザの口が位置するところの周辺に設けてもよい。

10

【0147】

このような構成にすることで、ユーザが意図的に発声したときの音を効率良く取得することが可能となる。

【0148】

また、接話用マイクロフォン961はユーザがファインダ150を覗いたとき、ユーザの口が位置するところに指向性を有するものでもよい。

【0149】

このような構成にすることで、ユーザがファインダ150の覗き方を変えた場合であっても、接話用マイクロフォン961を設ける位置を変えることなく、ユーザが意図的に発声したときの音を効率良く取得することが可能となる。

20

【0150】

また、非接話用マイクロフォン962は、ユーザが手でデジタルカメラ900を持ち、表示ディスプレイ151を見ながら発声する場合等、接話をしていない状態において好適に音を取得することが可能なマイクロフォンである。

【0151】

これは、ユーザが手でデジタルカメラ900を持ち、表示ディスプレイ151を見ながら発声する場合、腕の長さ程度の距離だけ離間した音源から出力された音を好適に取得できるマイクロフォンが適しているからである。

【0152】

また、デジタルカメラとユーザはほぼ正対するので、カメラ背面方向の指向性に優れたマイクロフォンを使用するのが望ましい。

30

【0153】

（外観の説明）

図10は、本実施形態に係るデジタルカメラ900の外観を示す図である。

【0154】

尚、図10（A）はデジタルカメラ900の前面の外観、図9（B）はデジタルカメラ900の背面の外観である。

【0155】

図9と共に通の要素には同一の符号を付し、その説明を省略する。

【0156】

40

（接話用マイクロフォン961と非接話用マイクロフォン962を切替える処理を示すフローチャートの説明）

図11は、本実施形態に係るデジタルカメラにおいて、表示モードに基づいて、音声入力の処理を設定する流れを説明するフローチャートである。

【0157】

ステップS1101において、制御部101は、音声処理操作ボタン141が押下されたか否か判断し、音声処理操作ボタン141が押下された場合、ステップS1102の処理を実行する。

【0158】

一方、音声処理操作ボタン141が押下されなかった場合、音声処理操作ボタン141

50

が押下されるまで待機する。

**【0159】**

ステップS1102において、制御部101は、表示ディスプレイ151に画像や撮像条件が表示されているか否か判断し、表示ディスプレイ151に画像や撮像条件が表示されている場合、ステップS1103の処理を実行する。

**【0160】**

即ち、表示部105の表示モードが第2の表示モードまたは第3の表示モードである場合、ステップS1103の処理を実行する。

**【0161】**

ステップS1103において、制御部101は、非接話用マイクロフォン962を用いて集音するように設定する。  
10

**【0162】**

一方、表示ディスプレイ151に画像や撮像条件が表示されていない場合、ステップS1104の処理を実行する。

**【0163】**

即ち、表示部105の表示モードが第1の表示モードである場合、ステップS1104の処理を実行する。

**【0164】**

ステップS1104において、制御部101は、接話用マイクロフォン961を用いて集音するように設定する。  
20

**【0165】**

次にステップS1105において、音声入力部106は音を取得する処理を開始する。

**【0166】**

(第5の実施形態)

本実施形態は、カメラ上に複数のマイクロフォンを配置し、マイクロフォンアレイを構成し、音声処理部106で音源推定や雑音抑圧処理を行って入力された音声の音質を向上させることを特徴とする。

**【0167】**

また、実施形態4において、接話用マイクロフォン、非接話用マイクロフォンにマイクロフォンアレイを用いても良い。  
30

**【0168】**

しかしながら、一般にマイクロフォンアレイによる各種処理は、処理量が増加するため、撮像装置の処理能力が乏しい場合は、入力音質が劣化する非接話用の音声入力時のみにマイクロフォンアレイを適用するよう構成してもよい。

**【0169】**

即ち、表示部150の表示モードが第3の表示モードの場合、即ちユーザがファインダを覗きながら操作する姿勢になる場合は、ユーザとカメラのマイクロフォンが接近することから比較的良好な音声入力環境になる。

**【0170】**

そこで、接話用の単一のマイクロフォンによる音声入力に設定する。  
40

**【0171】**

一方、表示部150の表示モードが第1の表示モードまたは第2の表示モードになる場合は、非接話用の設定、すなわちマイクロフォンアレイを適用した音声入力に設定する。

**【0172】**

このように構成すると、良好な音声入力環境を保ちながら、撮像装置における処理の負荷を軽減することができるという効果がある。

**【0173】**

(その他の実施形態)

また、本発明の目的は、以下の処理を実行することによって達成される。

**【0174】**

即ち、上述した実施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、システム或いは装置に供給し、そのシステム或いは装置のコンピュータ（またはC P UやM P U等）が記憶媒体に格納されたプログラムコードを読み出す処理である。

**【0175】**

この場合、記憶媒体から読み出されたプログラムコード自身が上述した実施の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及び該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。

**【0176】**

また、プログラムコードを供給するための記憶媒体としては、次のものを用いることができる。 10

**【0177】**

例えば、フレキシブルディスク、ハードディスク、光磁気ディスク、C D - R O M、C D - R、C D - R W、D V D - R O M、D V D - R A M、D V D - R W、D V D + R W、磁気テープ、不揮発性のメモリカード、R O M等である。

**【0178】**

または、ネットワークを介してプログラムコードをダウンロードしてもよい。

**【0179】**

また、コンピュータが読み出したプログラムコードを実行することにより、上記実施の形態の機能が実現される場合も本発明に含まれる。 20

**【0180】**

加えて、そのプログラムコードの指示に基づき、コンピュータ上で稼動しているO S（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。

**【0181】**

更に、前述した実施形態の機能が以下の処理によって実現される場合も本発明に含まれる。

**【0182】**

即ち、記憶媒体から読み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き込まれる。 30

**【0183】**

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C P U等が実際の処理の一部または全部を行う場合である。

**【図面の簡単な説明】**

**【0184】**

**【図1】**第1の実施形態に係る撮像装置の一例であるデジタルカメラを示す機能プロック図である。

**【図2】**第1の実施形態に係るデジタルカメラにおいて、音声入力の設定をする流れを説明するフローチャートである。 40

**【図3】**ユーザデジタルカメラを操作する際にとる姿勢を示す図である。

**【図4】**ユーザデジタルカメラを操作する際の音源の位置を示す図である。

**【図5】**第1の実施形態に係るデジタルカメラにおいて、音声処理の設定をする流れを説明するフローチャートである。

**【図6】**音声入力の設定をする流れの変形例を説明するフローチャートである。

**【図7】**第3の実施形態に係るデジタルカメラにおいて、認識モードを切替える処理を説明するフローチャート。

**【図8】**音声入力された制御コマンドに応じて実行する処理を示すデータテーブルを表す図である。

**【図9】**第4の実施形態に係る撮像装置の一例であるデジタルカメラを示す機能プロック 50

図である。

【図10】第4の実施形態に係るデジタルカメラの外観を示す図である。

【図11】第4の実施形態に係るデジタルカメラにおいて、音声入力を行うマイクロフォンを決定する流れを説明するフローチャートである。

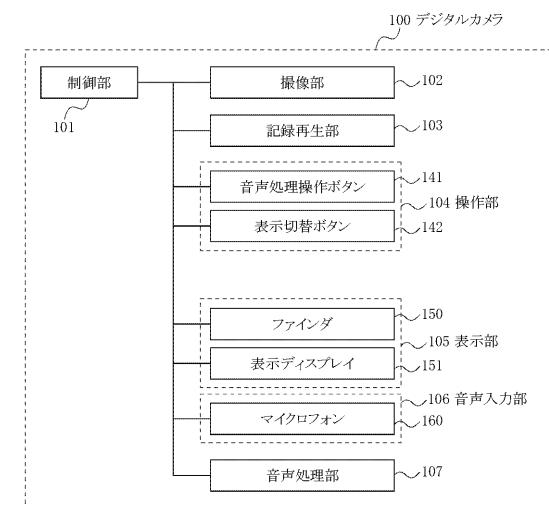
【符号の説明】

【0185】

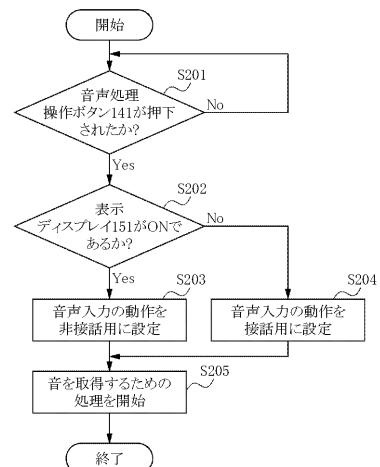
- 101 制御部
- 102 撮像部
- 103 記録再生部
- 104 操作部
- 105 表示部
- 106 音声入力部
- 107 音声処理部

10

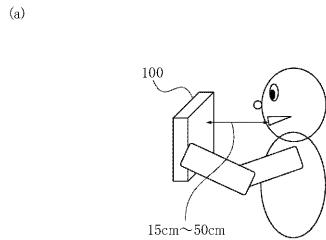
【図1】



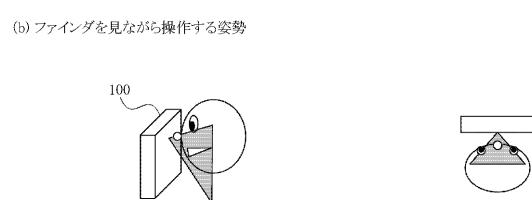
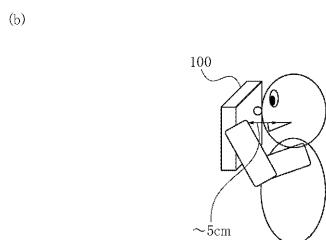
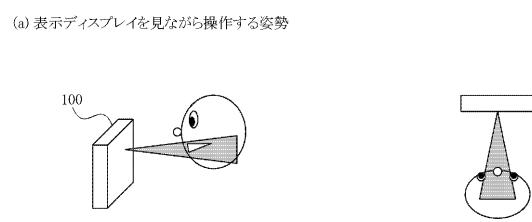
【図2】



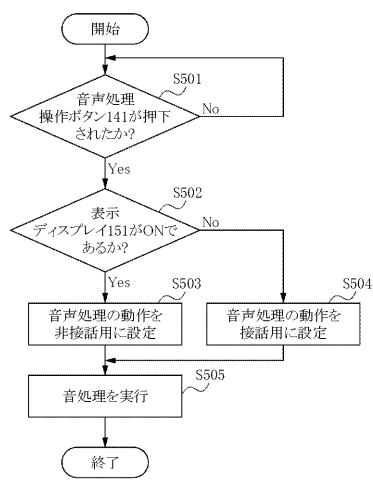
【図3】



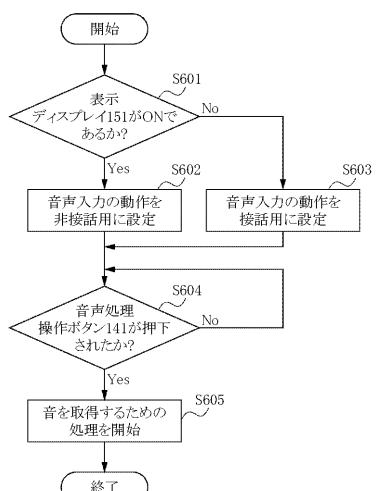
【図4】



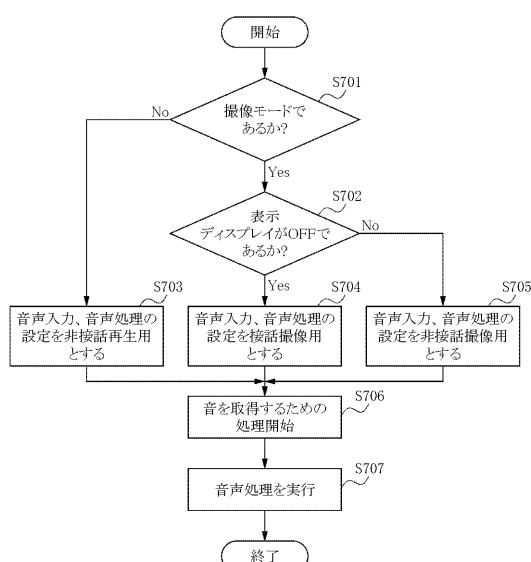
【図5】



【図6】



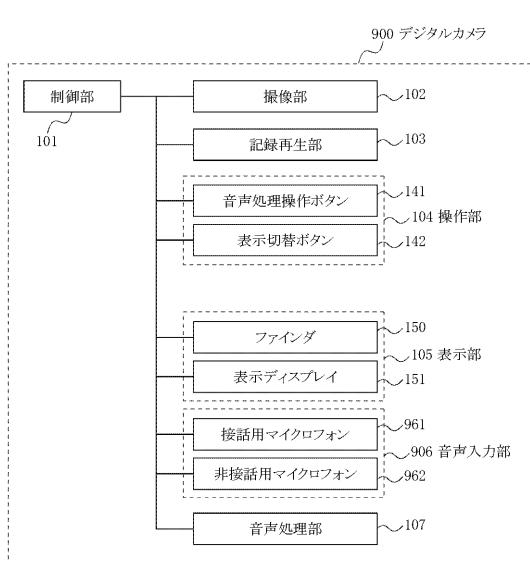
【図7】



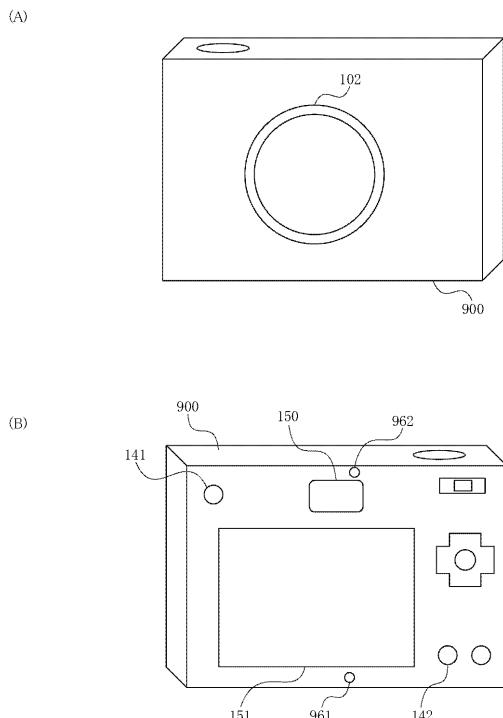
【図8】

	ディスプレイOFF+撮像モード (接話撮像)	ディスプレイON+撮像モード (非接話撮像)	ディスプレイON+再生モード (非接話再生)
「ストロボ自動発光」	ストロボ自動発光	—	—
「ストロボ発光」	ストロボ発光	—	—
「ストロボ発光禁止」	ストロボ発光禁止	—	—
「露出補正_明るく」	露出補正+1	—	—
「露出補正_暗く」	露出補正-1	—	—
「露出補正_標準」	露出補正±1	—	—
「露出補正_暗く」	露出補正-1	—	—
「通常撮影」	通常撮影	—	—
「人物撮影」	人物撮影	—	—
「風景撮影」	風景撮影	—	—
「接写撮影」	接写撮影	—	—
「夜景撮影」	夜景撮影	—	—
「記録画質L」	記録画質L	—	—
「記録画質M」	記録画質M	—	—
「記録画質S」	記録画質S	—	—
「記録画質RAW」	記録画質RAW	—	—
「一枚撮影」	一枚撮影	—	—
「連続撮影」	連続撮影	—	—
「セルフタイマー」	セルフタイマー	—	—
「ディスプレイON」	ディスプレイON+撮像モード	—	ディスプレイON+再生モード
「ディスプレイOFF」	—	ディスプレイOFF+撮像モード	ディスプレイON+再生モード
「撮像モード」	—	—	—
「再生モード」	ディスプレイON+再生モード	ディスプレイON+再生モード	—
「撮像」	撮像	撮像	—
「ズーム」	ズーム(光学)	ズーム(光学またはデジタル)	—
「ズーム」	ズーム(光学)	ズーム(光学またはデジタル)	—
「表示を明るく」	ファインダの明るさ+1	ディスプレイの明るさ+1	ディスプレイの明るさ+1
「表示を暗く」	ファインダの明るさ-1	ディスプレイの明るさ-1	ディスプレイの明るさ-1
「次の画像」	—	—	次の画像を表示
「前の画像」	—	—	前の画像を表示
「スライドショー」	—	—	スライドショー
「再生」	—	—	再生
「一時停止」	—	—	一時停止
「停止」	—	—	停止

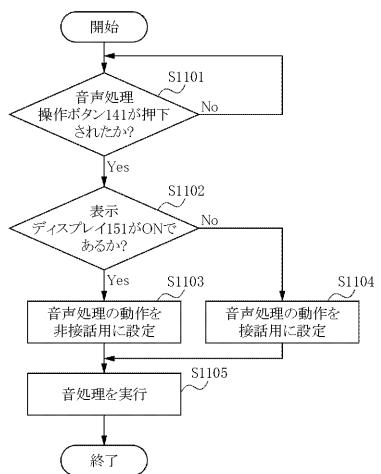
【図9】



【図10】



【図11】



---

フロントページの続き

(56)参考文献 特開平03-174132(JP,A)  
特開2005-024792(JP,A)  
特開2003-309746(JP,A)

(58)調査した分野(Int.Cl., DB名)

H04N 5/222 - 5/257  
G02B 17/00 , 17/26 - 17/34  
G02B 17/38 - 17/46  
G10L 15/00 - 17/00