

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4975813号
(P4975813)

(45) 発行日 平成24年7月11日(2012.7.11)

(24) 登録日 平成24年4月20日(2012.4.20)

(51) Int.Cl.	F I
HO4N 5/225 (2006.01)	HO4N 5/225 F
G1OL 15/06 (2006.01)	G1OL 15/06 300H
G1OL 15/00 (2006.01)	G1OL 15/00 200G
HO4N 5/232 (2006.01)	G1OL 15/06 200B
GO6F 3/16 (2006.01)	HO4N 5/232 B

請求項の数 9 (全 16 頁) 最終頁に続く

(21) 出願番号	特願2009-517461 (P2009-517461)	(73) 特許権者	502087507
(86) (22) 出願日	平成18年12月27日(2006.12.27)		ソニーモバイルコミュニケーションズ、
(65) 公表番号	特表2009-543396 (P2009-543396A)		エービー
(43) 公表日	平成21年12月3日(2009.12.3)		スウェーデン国、221 88 ルンド
(86) 国際出願番号	PCT/IB2006/055039	(74) 代理人	100076428
(87) 国際公開番号	W02008/004037		弁理士 大塚 康徳
(87) 国際公開日	平成20年1月10日(2008.1.10)	(74) 代理人	100112508
審査請求日	平成21年2月24日(2009.2.24)		弁理士 高柳 司郎
(31) 優先権主張番号	11/427, 807	(74) 代理人	100115071
(32) 優先日	平成18年6月30日(2006.6.30)		弁理士 大塚 康弘
(33) 優先権主張国	米国 (US)	(74) 代理人	100116894
			弁理士 木村 秀二
		(72) 発明者	グスタフソン、 マッツ
			スウェーデン国 ナッカ エス-131
			33, ケルバックスヴェーゲン 4
			最終頁に続く

(54) 【発明の名称】 音声による遠隔制御

(57) 【特許請求の範囲】

【請求項1】

少なくとも1つの撮像デバイスと、メモリを有する装置であって、
ユーザ入力にตอบสนองして、複数の制御動作及び、前記複数の制御動作の各々に関連付けられた音声コマンドを表示するように構成されたディスプレイと、

ロジックであって、

ユーザから音声入力を受信し、

前記音声入力を識別し、

前記識別された音声入力に基づいて、前記少なくとも1つの撮像デバイスに関する制御動作を実行し、

前記少なくとも1つの撮像デバイスに関する第1の制御動作に対応する少なくとも1つの単語又は語句を特定するテキストを、前記ユーザが前記装置に入力できるように構成されたユーザインタフェースを提供し、

前記テキストを前記第1の制御動作に対応するものとして前記メモリに保存するように構成されたロジックとを有し、

前記第1の制御動作が、写真撮影、ビデオ記録モードの起動、オートフォーカス機能の制御、ズームレンズの制御、又は解像度モードの制御の少なくとも1つを含むことを特徴とする装置。

【請求項2】

前記ロジックがさらに、

複数の制御動作の各々に対応する少なくとも1つの単語をユーザが提供することを可能にし、

前記複数の制御動作の各々に対応する前記少なくとも1つの単語を前記メモリに保存するように構成されることを特徴とする請求項1記載の装置。

【請求項3】

前記ロジックがさらに、

前記複数の制御動作の1つである第1の制御動作に関連付けられた第1の単語又は語句を識別する音声認識を実行し、

前記識別された第1の単語又は語句が正しいかをユーザに確認し、

前記識別された第1の単語又は語句が正しいと確認された場合、前記識別された第1の単語又は語句を前記第1の制御動作に対応するものとして前記メモリに保存するように構成されることを特徴とする請求項2記載の装置。

10

【請求項4】

セルラ電話機であることを特徴とする請求項1記載の装置。

【請求項5】

少なくとも1つの撮像デバイスと、ディスプレイと、メモリとを含んだ移動端末における方法であって、

ユーザインタフェースであって、

前記少なくとも1つの撮像デバイスに関する複数の制御動作を前記ディスプレイに出力し、

20

前記複数の制御動作の1つである第1の制御動作の選択をユーザから受信し、

前記ディスプレイを通じて、前記第1の制御動作に関する複数の選択肢を提供し、

前記第1の制御動作に対応する音声コマンドとして用いられる第1の単語又は語句を、前記複数の選択肢の選択を通じてユーザから受信し、

前記少なくとも1つの撮像デバイスに関する第2の制御動作に対応する少なくとも第2の単語又は語句を特定するテキストを前記ユーザが前記装置に入力できるように構成されたユーザインタフェース、を提供する工程と、

前記テキストを前記第2の制御動作に対応するものとして前記メモリに保存する工程と、

ユーザから音声入力を受信する工程と、

30

前記音声入力を前記第1の単語又は語句又は第2の単語又は語句に対応するものとして識別する工程と、

前記識別された音声入力に基づいて、前記少なくとも1つの撮像デバイスに関する前記第1の制御動作又は前記第2の制御動作を実行する工程とを有することを特徴とする方法。

【請求項6】

前記第1の制御動作及び前記第2の制御動作の各々が、写真撮影、ビデオ記録モードの開始、解像度モードの設定、フラッシュの起動、ズームレンズの制御、又はオートフォーカス機能の制御の少なくとも1つを含むことを特徴とする請求項5記載の方法。

【請求項7】

40

前記音声入力を識別する工程が、

音声認識ソフトウェアを用いて前記音声入力を識別する工程を有し、

前記音声認識ソフトウェアを用いて前記音声入力を識別する工程が、

前記音声入力を、保存されている、前記ユーザによって与えられた複数の音声信号と比較する工程と、

前記音声入力の前記保存されている音声信号の1つと合致するか判定する工程とを有することを特徴とする請求項5記載の方法。

【請求項8】

前記複数の制御動作の各々に関連付けられた音声入力を提供しようユーザに指示する工程と、

50

前記指示する工程にตอบสนองして、前記ユーザから、前記複数の制御動作の各々に関連付けられた少なくとも1つの単語を受信する工程と、

前記複数の制御動作の各々に関連付けられた前記少なくとも1つの単語を保存する工程とをさらに有することを特徴とする請求項5記載の方法。

【請求項9】

第3の制御動作に関連付けされた少なくとも第1の単語を識別するための音声認識を実行する工程と、

前記第3の制御動作に関連付けられた前記識別された第1の単語が正しいかユーザに確認する工程と、

前記識別された第1の単語が正しいと確認された場合、前記第3の制御動作に関連付けられた前記第1の単語を保存する工程とをさらに有することを特徴とする請求項8記載の方法。

10

【発明の詳細な説明】

【技術分野】

【0001】

本発明は全体として制御動作の実行に関し、特に、音声認識を用いた、撮像に関する制御動作の実行に関する。

【背景技術】

【0002】

携帯電話機のような通信機器は、ますます多機能になってきている。例えば、今日の多くの携帯電話機には、テキストメッセージの送信/受信、音楽再生、ビデオゲームの実行など、ユーザが電話の発信や受信以外のことを行えるようなアプリケーションが含まれている。また、カメラを有する携帯電話機も多く、ユーザはカメラによって写真撮影やビデオクリップの記録が可能である。結果として、携帯電話機はますます日常生活に係わるようになってきている。

20

US2005118990は携帯電話機と、オーディオ又は音声制御を通じて制御可能なデジタルカメラの組み合わせに関する。携帯電話機は可聴入力を検出するための少なくとも1つのマイクを含む。検出された可聴入力は電気信号に変換され、プロセッサに送信される。プロセッサは、検出された可聴入力をプロセッサに処理させるソフトウェアを格納する記憶媒体へアクセス可能である。プロセッサは検出された可聴入力をカメラコマンドのデータベースと比較し、検出された可聴入力データベース中のカメラコマンドの1つと合致するか判定する。合致した場合、プロセッサは、求められた機能を実行するため、そのカメラコマンドに関連付けられた命令群を実行する。

30

US 5027249には、音声認識動作によって写真を撮影するカメラが記載されている。カメラはマイクを有し、操作者が発声したコマンド語はマイクを通じてデジタルデータの形式でコンピュータに登録される。コンピュータは、登録されているデータに従って同一の発声されたコマンド語を特定し、カメラの意図された機能を実行する。カメラは、操作者が自分がカメラに行った操作を確認できるように、発声されたコマンド語の登録及び認識に表示装置を用いるコンピュータの機能を示している。

音声コマンドシステムを選択的に学習及び稼働モードの一方にするスイッチを含む音声コマンドを有するカメラが、US 4389109に記載されている。学習モードの間、様々なコマンド音声はマイクに発声され、それぞれがカメラ内の指定されたセクションに登録される。稼働モードの間、発声されたコマンド音声は登録されたコマンド音声の各々と比較され、特定されたセクションに応じて制御信号を生成するため、相関の最も大きなコマンド音声と認識される。

40

【発明の概要】

【0003】

1つの見地によれば、少なくとも1つの撮像デバイスおよびロジックを有する装置が提供される。前記ロジックは、ユーザから音声入力を受信し、前記音声入力を識別し、前記識別した音声入力に基づいて前記少なくとも1つの撮像デバイスに関する制御動作を実行

50

するように構成される。

【0004】

加えて、前記少なくとも1つの撮像デバイスはカメラを有することができ、前記制御動作は前記カメラを用いた写真撮影を含む。

【0005】

さらに、前記少なくとも1つの撮像デバイスはビデオ録画デバイスを有することができ、前記制御動作は前記ビデオ録画デバイスに関するビデオ記録モードの始動を含む。

【0006】

さらに、前記制御動作は、オートフォーカス機能の制御、ズームレンズの制御、又は解像度モードの制御を含むことができる。

10

【0007】

さらに、前記音声入力を識別する際、前記ロジックは音声認識ソフトウェアを用いて前記音声入力を識別するように構成される。

【0008】

さらに、装置は、ユーザ入力にตอบสนองして、複数の制御動作及び、前記複数の制御動作の各々に関連付けられた複数の制御動作を表示するように構成されたディスプレイを有することができる。

【0009】

さらに、装置はメモリを有することができ、前記ロジックはさらに、複数の制御動作の各々に対応する少なくとも1つの単語をユーザが与えることを可能にするように構成されてもよい。前記ロジックはさらに、前記複数の制御動作の各々に対応する前記少なくとも1つの単語を前記メモリに保存するように構成されてもよい。

20

【0010】

加えて、前記ロジックはさらに、前記複数の制御動作の1つである第1の制御動作に関連付けられた第1の単語又は語句を識別するための音声認識を実行し、前記識別された第1の単語又は語句が正しいかユーザに確認し、前記識別された第1の単語又は語句が正しいと確認された場合、前記確認された第1の単語又は語句を、前記第1の制御動作に対応するものとして前記メモリに保存するように構成されてもよい。

【0011】

さらに、装置はメモリを有することができ、前記ロジックはさらに、ユーザが前記装置に、前記少なくとも1つの撮像デバイスに関する第1の制御動作に対応する少なくとも1つの単語又は語句を特定するテキストを入力できるように構成されたユーザインタフェースを提供するとともに、前記テキストを前記第1の制御動作に対応するものとして前記メモリに保存するように構成されてもよい。

30

【0012】

さらに、前記装置は携帯電話機を有してもよい。

【0013】

別の見地によれば、少なくとも1つの撮像デバイスを含んだ移動端末において方法が実行される。前記方法は、ユーザから音声入力を受信する工程と、前記音声入力を識別する工程と、前記識別された音声入力に基づいて、前記少なくとも1つの撮像デバイスに関する制御動作を実行する工程とを含む。

40

【0014】

さらに、前記制御動作は、写真撮影、ビデオ記録モードの開始、解像度モードの設定、フラッシュの起動、ズームレンズの制御、又はオートフォーカス機能の制御の少なくとも1つを含むことができる。

【0015】

さらに、前記音声入力を識別する工程は、音声認識ソフトウェアを用いて前記音声入力を識別する工程を有する。

【0016】

さらに、音声認識ソフトウェアを用いて前記音声入力を識別する工程は、前記音声入力

50

と、保存されている、ユーザによって与えられた複数の音声信号とを比較する工程と、前記音声入力の前記保存されている音声信号の1つと合致するか判定する工程とを有する。

【0017】

さらに、前記方法は、複数の制御動作の各々に関連付けられた音声入力を提供するようにユーザに指示する工程と、前記指示する工程にตอบสนองして、前記ユーザから、前記複数の制御動作の各々に関連付けられた少なくとも1つの単語を受信する工程と、前記複数の制御動作の各々に関連付けられた前記少なくとも1つの単語を保存する工程とをさらに有してもよい。

【0018】

さらに、前記方法は、前記複数の制御動作の1つである第1の制御動作に関連付けられた少なくとも第1の単語を識別するための音声認識を実行する工程と、前記第1の制御動作に関連付けられた前記識別された第1の単語が正しいかユーザに確認し、前記識別された第1の単語が正しいと確認された場合、前記第1の制御動作に関連付けられた前記第1の単語を保存する工程とをさらに有してもよい。

10

【0019】

さらに別の見地によれば、装置が提供される。装置は、写真又はビデオの少なくとも一方を撮影する手段と、ユーザから音声入力を受信する手段と、前記音声入力に基づいて、前記写真又はビデオの少なくとも一方を撮影する手段に関連付けられた制御動作を実行する手段とを有する。

【0020】

20

さらに、前記装置は、複数の制御動作と、前記複数の制御動作に対応する複数の音声コマンドとを表示する手段とをさらに有してもよい。

【0021】

さらに別の見地によれば、複数の命令が記録されたコンピュータ読み取り可能な媒体が提供される。前記命令は、少なくとも1つのプロセッサが実行した際、前記少なくとも1つのプロセッサに、ユーザから音声入力を受信させ、前記音声入力を識別させ、そして、前記識別された音声入力に基づいて、撮像に関する制御機能を実行させる。

【0022】

さらに、前記撮像に関する制御動作は、カメラを用いた写真撮影、ビデオ録画デバイスに関連付けられたビデオ記録モードの開始、解像度モードの設定、フラッシュの起動、ズームレンズの制御、又はオートフォーカス機能の制御の少なくとも1つを含むことができる。

30

【0023】

さらに、前記命令は、前記プロセッサに、ユーザ入力にตอบสนองして、撮像に関する複数の制御動作と前記撮像に関する複数の制御動作の各々に対応する音声コマンドとを表示させる。

【0024】

さらに、前記命令は、前記プロセッサに、撮像に関する複数の制御動作の各々に対応する単語又は語句をユーザが選択可能とさせる。

【0025】

40

以下の詳細な説明から、本発明の他の特徴及び利点は本技術分野の当業者に明らかになるであろう。ここで示され、また、説明される実施形態は、本発明の実施を考慮したベストモードの実例を提供する。本発明はその範囲内で、様々な明らかな観点から変更することが可能である。従って、図面は実際には例示的なものであって、限定的なものとは見なされるべきものではない。

【図面の簡単な説明】

【0026】

【図1A】、

【図1B】本発明に従った方法及びシステムを実施できるであろう例示的な移動端末の図である。

50

【図2】本発明に従った実施に係る、図1A及び図1Bの移動端末の例示的なブロック図である。

【図3】本発明に従った実施に係る図2の移動端末において実施される構成要素の例示的な機能ブロック図である。

【図4】本発明に係る図2の移動端末の構成に関する例示的な処理を示すフローチャートである。

【図5】図2の移動端末に格納され及び/またはユーザに表示されてよい例示的なテーブルを示す図である。

【図6】本発明に係る例示的な処理を示すフローチャートである。

【発明を実施するための形態】

10

【0027】

添付図面を通じて、同様の参照番号の指定は同様の要素を表している。

以下、添付図面を参照して本発明を詳細に説明する。異なる図面における同一の参照番号は、同一又は類似の要素を示している。また、以下の詳細な説明は、本発明を限定しない。代わりに、本発明の範囲は、添付の請求項とその等価物によって規定される。図1A及び図1Bは、本発明に従った方法及びシステムを実施できるであろう例示的な移動端末の正面図及び背面図を示している。ここでは、本発明を移動端末に関して説明する。また、本明細書において、“移動端末”とは、セルラ無線電話機（複数行ディスプレイの有無を問わず）、セルラ無線電話機にデータ処理機能、ファクシミリ機能及びデータ通信機能とを組み合わせたものであってよいパーソナル通信システム（PCS）端末、無線電話、
20
ポケットベル、インターネット/イントラネットアクセス、ウェブブラウザ、スケジュール帳(organizer)、カレンダー及び/又はGPS受信機を含みうる携帯情報端末（PDA）、及び、従前のラップトップ及び/又はパームトップ受信機、又は無線電話送受信器を含む他の装置を含みうる。移動端末はまた、“広範囲(pervasive)コンピューティング”装置とも呼ばれうる。

20

【0028】

なお、本発明は、通話の発信および着信に関する通信機能を持たない他の機器においても実施可能であることを理解すべきである。例えば、本発明はいかなる撮像装置においても実施されうる。本明細書において、“撮像装置”とは、カメラ及び/又はビデオ録画機器のような、静止画及び/又は動画を撮影する任意の機器を含みうるものとして用いる。
30

30

【0029】

図1Aにおいて、移動端末100は、筐体110、レンズ120、及びシャッターボタン130を含むことができる。筐体110は移動端末100の構成要素を風雨などから保護することができる。レンズ120は光の焦点を合わせることができ、また複数のレンズ要素を含んでよい。レンズカバー（図示せず）は、ユーザに撮影を許すためにレンズ120を露出させ、カメラが使用されていない際にはレンズを保護するように制御されてよい。シャッターボタン130は撮影時にユーザに押下されてよい。

【0030】

図1Bは、端末100の背面を示す。図1Bにおいて、移動端末100は表示画面140及びユーザ制御エリア150を含んでいる。表示画面140は、画像液晶ディスプレイ(LCD)又は他のタイプの表示画面であってよく、ユーザが写真撮影及び/又はビデオクリップの記録を行う際に、撮影される画像を確認することを可能にする。一部の実施形態において、レンズ120は移動端末100の表示画面140と同じ側に位置していてもよい。移動端末100はまた、写真撮影及び/又はビデオクリップの記録に関する他の要素/部品（図示せず）を含んでいてもよい。例えば、移動端末100はズームレンズ、フラッシュ、及び、写真及び/又はビデオ撮影を助ける他の要素を含むことができる。ユーザ制御エリア150は、移動端末100を用いた通話に関する複数のコントロールを含んでよい。
40

40

【0031】

例えば、ユーザ制御エリア150はダイヤルボタン、通話切断ボタンなどを含むことが
50

50

できる。ユーザ制御エリア 150 はさらに、以下に詳細を説明するような、ユーザが機能（例えば、移動端末 100 についての音声作動式リモート制御機能）の選択に関するメニューを見ることを可能にするメニューボタンを含むことができる。

【0032】

図 2 は、本発明に従った例示的な実施形態に係る移動端末 100 の図である。移動端末 100 は、バス 210、処理ロジック 220、メモリ 230、入力デバイス 240、出力デバイス 250、撮像デバイス 260、及び通信インタフェース 270 を含んでよい。バス 210 は移動端末 100 の構成要素間の通信を可能にする。本技術分野の当業者は、移動端末 100 が様々な方法で構成可能であり、他の、あるいは異なる要素を含みうることを理解するであろう。例えば、移動端末 100 は 1 つ又は複数の電源（図示せず）を含んでよい。移動端末 100 はさらに、データを処理するための変調器、復調器、符号化器、復号化器、等を含みうる。

10

【0033】

処理ロジック 220 はプロセッサ、マイクロプロセッサ、特定用途向け集積回路(ASIC)、フィールドプログラマブルゲートアレイ(FPGA)、等を含むことができる。処理ロジック 220 は、移動端末 100 の動作を制御するため、ソフトウェア命令/プログラム、又はデータ構造を実行することができる。

【0034】

メモリ 230 は、処理ロジック 220 が実行するための情報及び命令を格納するランダムアクセスメモリ(RAM)又は他の形式の動的ストレージデバイス；処理ロジック 220 によって用いるための静的な情報及び命令を格納する読み出し専用メモリ(ROM)又は他の形式の静的ストレージデバイス；情報及び命令を格納するためのフラッシュメモリ（例えば、電氣的消去可能プログラマブルメモリ(EEPROM)）デバイス；及び/又は、他の形式の磁気又は光学記録媒体と、媒体に対応するドライブとを含むことができる。メモリ 230 はさらに、処理ロジック 220 が命令を実行中に一時変数又は他の中間情報を格納するためにも利用されうる。処理ロジック 220 によって用いられる命令はさらに、又は代わりに、処理ロジック 220 からアクセス可能な他の形式のコンピュータ読み取り可能媒体に格納されてもよい。

20

【0035】

入力デバイス 240 は、マイク、キーボード、キーパッド、マウス、ペン、音声認識、及び/又は生体を利用する機構など、操作者が移動端末 100 に情報を入力することを可能にする任意の機構を含むことができる。入力デバイス 240 はさらに、ユーザが出力デバイス 250 を介してオプションのメニューを受信することを可能にする、（ユーザ制御エリア 150 におけるボタンのような）1 つ又は複数のボタンを含むことができる。このメニューは、移動端末 100 が実行するアプリケーションに関する様々な機能又はモードの選択をユーザに許すものであってよい。あるいは、入力デバイス 240 は、ユーザが移動端末 100 の、（カメラ 260 に関する、音声作動型リモート制御機能に関するモードのような）特定のモードを起動することを可能にする 1 つ又は複数のボタンを制御エリア 150 に含んでもよい。

30

【0036】

出力デバイス 250 はディスプレイ、プリンタ、1 つ又は複数のスピーカ等の、ユーザに情報を出力する 1 つ又は複数の一般的な機構を含んでよい。出力デバイス 250 はさらに、入来呼を受信した際に移動端末 100 を振動させる振動機構を含むことができる。

40

【0037】

撮像デバイス 260 は、レンズ 120、シャッターボタン 130 及び他の構成要素といった、移動端末 100 での写真撮影を可能にするための一般的なカメラ部品を含むことができる。撮像デバイス 260 はさらに、撮像デバイス 260 がビデオクリップのような動画を記録することを可能にするための構成要素を含むことができる。撮像デバイス 260 は写真/ビデオクリップを、例えばメモリ 230 又は他のメモリに保存することができる。上述の通り、撮像デバイス 260 はズームレンズ、フラッシュ等、他の要素をさらに含

50

むことができる。撮像デバイス260はさらに、2つのモードのうちの1つで動作することができる。第1のモードにおいて、移動端末100のユーザは、シャッターボタン130を押下することによって撮影し、手動でズームレンズを操作し、及び/又は移動端末100上の1つ又は複数のボタンを押下することによりビデオモードへの切替を行うことができる。第2のモードにおいて、移動端末100のユーザは、以下に詳細を説明するように、音声コマンドを用いて、撮像デバイス260による写真/ビデオを撮影することができる。図2において、撮像デバイス260は1つのデバイスとして示されている。一部の実施形態において、撮像デバイス260は、静止画撮影用の独立したカメラ関連要素/部品、及び動画(例えばビデオ)を撮影、記録するためのビデオ記録に関連する独立した要素/部品を含むことができる。

10

【0038】

通信インタフェース270は、移動端末100が他の装置及び/又はシステムと通信することを可能にする、任意の送受信器様の機構を含むことができる。例えば、通信インタフェース270はモデム又はLANへのイーサネットインタフェースを含むことができる。

【0039】

通信インタフェース270は、無線ネットワークのようなネットワークを介して通信するための他の手段をさらに有することができる。例えば、通信インタフェース270は無線周波数(RF)データを伝送するため、1つ又は複数のRF送信器及び受信器、及び/又は送受信器を含んでもよい。通信インタフェース270は、RFデータのようなデータを送信/受信するための1つ又は複数のアンテナをさらに含むことができる。

20

【0040】

本発明に係る移動端末100は、撮像ロジック260のような、移動端末100上の様々な要素の制御に関する処理を実行することができる。移動端末100はそれらの動作を、メモリ230のようなコンピュータ読み取り可能な媒体に格納されている、命令シーケンスを処理ロジック220が実行することに応答して実行することができる。そのような命令は、他のコンピュータ読み取り可能な媒体から、例えば通信インタフェース270を介してメモリ230に読み込まれてもよい。コンピュータ読み取り可能な媒体は、1つ又は複数のメモリデバイス及び/又は搬送波を含みうる。代替実施形態においては、本発明に従った処理を実行するため、ソフトウェア命令に代えて、もしくはソフトウェア命令と組み合わせて、ハードウェア回路を用いることができる。従って、本発明に従った実施形態は、ハードウェア回路及びソフトウェアの特定の組み合わせには限定されない。

30

【0041】

図3は、本発明に従った実施形態に係る図2の移動端末100において(例えばメモリ230において)、実施される構成要素の例示的な機能ブロック図である。図3に示すように、メモリ230は音声認識プログラム310及び制御コマンド生成器320を含むことができる。音声認識プログラム310は、様々な単語及び/又は語句を識別し、認識した単語/語句に基づいて情報を出力する音声認識を実行可能な任意のタイプの音声認識ソフトウェア及び/又はハードウェアを含むことができる。音声認識プログラム310はさらに、様々な動作に関連付けられる単語/語句を与えるよう移動端末100のユーザに指示するユーザインタフェースを含むことができる(詳細は以下で説明する)。音声認識プログラム310はさらに、音声認識の実行に係る複数の規則又は文法を保存するルールデータベースを含んでもよい。これら規則/文法は、音声認識プログラム310が認識するように構成されている特定の単語/語句に基づくことができる。

40

【0042】

例えば、音声認識プログラム310は、“チーズ”、“クリック”、“撮影”、“ズームイン”、“広角”、“フォーカス”、“オートフォーカス通常”、“オートフォーカスマクロ”、“オートフォーカスオフ”、“バースト4”、“ビデオスタート”、“ビデオストップ”、“フラッシュ”、“高解像度”、“低解像度”、“マイクオン”、“マイクオフ”、といった様々な単語、及び、撮像デバイス260を用いた写真及び/又はビデオ

50

の撮影に関する他の単語 / 語句を認識するように構成されうる。以下に説明するように、これらの単語 / 語句は移動端末 100 のユーザによって与えられ、様々な制御動作に関連付けられて、移動端末 100 に保存されてよい。音声認識プログラム 310 は、精度の良い音声認識の一助としてこれら保存された複数の単語 / 語句を用いることができる。

【0043】

音声認識プログラム 310 はさらに、移動端末 100 のユーザが入力した単語 / 語句のような、様々な制御動作に関連付けることをユーザが希望する音声サンプルを保存してもよい。音声認識プログラム 310 は、音声認識処理の一助としてこれら保存された複数の音声サンプルを用いることができる。例えば、音声認識プログラム 310 は、ユーザが入力した単語又は語句を、保存されている音声サンプルと比較してもよい。受け付けた単語又は語句が保存されている音声サンプルの 1 つと一致した場合、音声認識プログラム 310 は保存されている音声サンプルに関連付けられている単語 / 語句が、ユーザが発生した単語 / 語句であると識別する。保存されている音声サンプルの使用は、音声認識処理の精度向上の一助となるであろう。

10

【0044】

いずれの場合も音声認識プログラム 310 は、音声入力を受信し、音声入力が音声認識プログラム 310 に保存されているいずれか 1 つの単語 / 語句に対応するかどうか判定することができる。そして、音声認識プログラム 310 は発声された単語 / 語句を特定する情報を、制御コマンド発生器 320 のような、移動端末 100 内の他のデバイス / アプリケーションへ転送することができる。

20

【0045】

制御コマンド発生器 320 は、音声認識プログラム 310 から情報を受信し、移動端末 100 が実行するアプリケーションを制御するためのコマンドを生成することができる。例えば、一実施形態において、制御コマンド発生器 320 は、撮像デバイス 260 及び / 又は、ズームレンズやフラッシュのような撮像デバイス 260 に付随する周辺機器によって写真及び / 又はビデオを撮影することに関連付けられた動作を制御するように構成されてよい。

【0046】

上述の通り、一部の実施形態において、音声認識プログラム 310 は、様々な制御コマンドに関連付けされる様々な単語 / 語句をユーザが与えることを可能にするユーザインタフェースを含むことができる。例えば、図 4 は、音声認識プログラム 310 及び / 又は制御コマンド発生器 320 が様々な制御動作を実行するために用いられる、カスタマイズされた音声コマンドの提供に関する例示的な処理を示している。処理は、(例えば移動端末 100 の出力デバイス 250 を介して) 提供されるメニューにユーザがアクセスすることによって開始されてよい。このメニューは、音声認識プログラム 310 によって認識されることになる様々な音声コマンドの入力をユーザに許可するオプションを含むことができる。ユーザがこのオプションを選択したとする。

30

【0047】

音声認識プログラム 310 は、移動端末 100 のユーザに、ユーザインタフェースを提供することができる(動作 410)。例示的な一実施形態において、ユーザインタフェースは、表示画面 140 を通じ、撮影、ズームイン、ズームアウト、オートフォーカス、撮影時フラッシュ点灯、ビデオモードへの移行、高解像度モードの使用、低解像度モードの使用、カメラのオン、カメラのオフ、マイクのオン、マイクのオフ等、複数のコマンドをユーザに表示することができる。

40

【0048】

ユーザが、撮影コマンドを選択したとする。この場合、ユーザインタフェースは、撮像デバイス 260 を用いた撮影のためのコマンドにユーザが対応付けたい単語又は語句を発声するよう、ユーザに指示することができる。音声認識プログラム 310 が与える指示は、出力デバイス 250 (例えばスピーカ)を介した音声指示出力、及び / 又は、表示画面 140 を介したテキスト指示出力であってよい。

50

【0049】

ここでは、ユーザが単語、“チーズ”を撮影のための音声コマンドに対応付けたいものとする。この場合、ユーザは単語、“チーズ”を発声することができる。音声認識プログラム310は、例えば入力デバイス240（例えばマイク）を通じて、この音声入力を受信することができる（動作420）。音声認識プログラム310は、音声認識プログラム310が正しく単語を識別したことを確実にするため、単語“チーズ”をユーザに対して発声し返してもよい。

【0050】

例えば、音声認識プログラム310は、「あなたは単語“チーズ”を撮影のためのコマンドとして選択しました。間違いありませんか？」といった音声出力を行ってもよい。音声認識プログラム310が単語/語句（この例では“チーズ”）を正しく識別している場合、ユーザは、イエスと応答することなどにより、音声認識プログラム310が識別した単語/語句を確認することができる。

10

【0051】

あるいは、音声認識プログラム310によって与えられる指示が、入力デバイス240（例えば英数字キーパッド）を用いて行うようなテキスト入力によって単語/語句を入力するようユーザに指示してもよい。

【0052】

いずれの場合も、音声認識プログラム310はユーザによって与えられた単語/語句を保存し、その単語/語句を撮影のためのコマンドに関連付ける（動作440）。複数の他のコマンド/動作について、同様の処理を継続してもよい。すなわち、ユーザは、所定量ズームイン、最大量ズームイン、所定量ズームアウト、最大量ズームアウト、オートフォーカス機能の使用、高速連写、撮影時フラッシュ点灯、ビデオ撮影モードの開始、ビデオ撮影モードの停止、写真撮影/保存に関する高解像度モードの起動、写真撮影/保存に関する低解像度モードの起動、カメラのオン、カメラのオフ、マイクのオン、マイクのオフ、等に対する単語/語句を与えることができる。いずれの場合も、ユーザは所望の動作を実行するための好みの単語又は語句を選択することができる。

20

【0053】

例えば、上述の通り、単語/語句は、写真撮影についての“チーズ”や、撮像デバイス260をビデオ記録モードで動作させるための“ビデオ”のように、1語であってよい。あるいは、ズームレンズを用いて被写体をズームインするための“最大ズーム(max zoom)”、カメラ260を用いて実現可能な最大メガピクセル数を用いてデジタル写真を撮影及び保存するよう撮像デバイス260に指示する“高解像度(high resolution)”、4枚の写真を高速連写するよう撮像デバイス260に指示する“バースト4(burst 4)”のように、単語/語句は複数の単語であってよい。このようにして、ユーザの好む音声コマンドを移動端末100に保存し、所望の制御動作に関連付けることができる。

30

【0054】

最初のコマンドがユーザから与えられた後で、ユーザは、移動端末100が保存している単語/語句及び対応するコマンドのリストを見ることができる。例えば、移動端末100はユーザが移動端末100に保存されている音声コマンドを閲覧することを可能にするメニューを含むことができる。ユーザは、例えばユーザ制御エリア150（図1B）のボタンを押下して、コマンドを閲覧することができる。ユーザのこの選択に応答して、移動端末100はコマンド及び対応する音声入力のリストを表示画面140を介して出力することができる。

40

【0055】

例えば、図5は、移動端末100に保存され、及び/又は表示画面140を介してユーザに表示される例示的なテーブルを示している。図5において、テーブル500はフィールド510に制御動作のリストを、対応する音声コマンドのリストをフィールド520に含むことができる。例えば、最初のエントリに図示されるように、カメラをオンするための音声コマンドは、“カメラオン”によるものであってよい。音声コマンドを変更したい

50

場合、ユーザはキーパッドコントロールデバイスのような入力デバイスを用いて特定のコマンドを選択し、新しいコマンドを与えることができる。例えば、ユーザが撮影用のコマンドを"チーズ"から"クリック"に変えたいとすると、ユーザはフィールド510の動作"撮影"又はフィールド520の音声コマンド"チーズ"を選択し、"チーズ"を置き換えるための新たな単語/語句を与えることができる。この例では、ユーザは"クリック"と発声(もしくは英数字キーパッドを用いて"クリック"と入力)し、音声認識プログラム310は"チーズ"を"クリック"に置き換え、"クリック"を撮影用のコマンドとして認識するであろう。このようにして、ユーザは特定の制御動作に対する特定の音声コマンドをいつでも変更することができる。

【0056】

一部の実施形態において、音声認識プログラム310は、個々の制御動作に用いることのできる選択肢(例えば単語/語句)のメニューを保存することができる。この場合、音声認識プログラム310は特定の制御動作に対する選択肢を表示することができ、ユーザはその特定の制御動作に対して使用したい1つの選択肢を選択することができる。そして、ユーザは選択した単語/語句を所望の制御動作を実行するための音声コマンドとして使用することができる。

【0057】

さらに、一部の実施形態においては、デフォルトの音声コマンドが音声認識プログラム310によって予め保存されてもよい。つまり、音声認識プログラム310は、ユーザが特定の単語/語句を入力する前に、特定の制御動作に対する様々な音声コマンドを保存してもよい。この場合、ユーザは音声コマンドを閲覧し、予め保存されている単語/語句を用いるか、音声コマンドに対応する予め保存されている単語/語句に対して1つ又は複数の変更を行うかを決定することができる。いずれの場合も、音声認識プログラム310は、ユーザが様々な制御動作を実行するための好みの単語/語句を選択することを可能にする。そして、以下に詳細に説明するように、移動端末100は、音声入力にตอบสนองして、求められる制御動作を実行することができる。

【0058】

図6は、撮像デバイス260に関する制御動作の実行に関する例示的な処理を説明するフローチャートである。ここで、移動端末100の電源はオンされており、移動端末100のユーザが撮像デバイス260を用いた写真撮影を希望しているものとする。例えば、移動端末100のユーザが、撮像デバイス260を用いて友人たちと一緒に自分の写真撮影したいとする。この場合、移動端末100のユーザは移動端末100をテーブルのようなどこかの面に置き、レンズ120の前で自分の周りに友人たちを並べ、単語"チーズ"と発声することができる。

【0059】

音声認識プログラム310は音声入力を受信し、ユーザが発声した特定の単語/語句を識別するために音声認識を実行することができる(動作610)。ここでは、音声認識プログラムが単語"チーズ"を識別したものとする。そして、音声認識プログラム310は、発声された単語/語句(本例では"チーズ")に対応するコマンドを特定することができる(動作620)。例えば、音声認識プログラム310はテーブル500にアクセスし、単語"チーズ"が撮像のための制御動作510に対応することを特定することができる。

【0060】

音声認識プログラム310は特定したコマンドを制御コマンド生成器320に転送することができる。制御コマンド生成器320は音声認識プログラム310から情報を受け取り、撮像デバイス260に写真撮影を行わせるための適切なコマンド又は命令を生成する(動作630)。例えば、制御コマンド生成器320は、撮像デバイス260に写真撮影を行うように命令することができる。この命令は、シャッターボタン130が押下された際に受信される従前の信号に代わることができる。そして、撮像デバイス260は写真撮影することができる(動作640)。このようにして、移動端末100のユーザは撮影

10

20

30

40

50

のためのコマンドを発声するだけで、移動端末100は写真撮影に必要な制御動作を実行する。これによりユーザは、所望の動作を実行させるためにシャッターボタン130を押下したり、他の1つ又は複数のボタンを押下したりすること無しに、撮像デバイス260に関する様々な制御動作を実行させることが可能になる。このような方法で音声認識プログラム310を用いることにより、タイマ機器を設定してから走ったりすることなく、ユーザが自分を含めた写真を撮影することを可能にすることができる。さらに、音声認識プログラム310を用いることで、シャッターボタン130を押下した際に移動端末100を不適切に動かしてしまうことを回避可能にすることができ、写真がピンぼけになったり、中心がずれたりすることを回避可能とすることができる。

【0061】

上述したように、制御コマンド生成器320は、ズームレンズ、オートフォーカス機能、フラッシュなどの、撮像デバイス260に付随する他の構成要素を制御するコマンド又は命令も生成することができる。いずれの場合も、制御コマンド生成器320は、求められる制御機能を実行するために撮像デバイス260に付随する様々な構成要素によって認識されるコマンド、コード、及び/又は命令を保存するように構成されてよい。

【0062】

別の例として、移動端末100に関連付けられたユーザがビデオ記録モードへの切替を希望しているとする、ユーザは”ビデオ開始”と発声するだけで、撮像デバイス260はビデオクリップ(例えば動画像)を記録することができるビデオ記録モードに切り替わるであろう。このような方法で音声認識プログラム310を用いることにより、ユーザが特定の被写体に集中していて忙しい場合の付加価値を提供することができる。例えば、ユーザがスポーツイベントでの活動を追っている場合、ユーザは移動端末100上の1つ又は複数のボタンを見て押下する必要なく、単にビデオモードへ切り替えるためのコマンドを発声すればよい。

【0063】

さらに別の例として、移動端末100のユーザが、撮像デバイス260を用いて実現できる最高解像度を用いた写真撮影を希望しているとする。この場合、ユーザは”最大解像度”と発声することができ、撮像デバイス260は自身を用いて実現可能な最大メガピクセルを用いて写真撮影を行うであろう。ここでも、このような方法で音声認識プログラム310を用いることで、撮像デバイス260に関する写真解像度モードを変更するために1つ又は複数のボタンを押下する必要がなく、ユーザは写真撮影したい被写体に集中することが可能になる。

【0064】

さらに別の状況において、移動端末100のユーザは複数の写真(例えば4枚の写真)を高速で連写することを希望するかもしれない。この場合、ユーザは”バースト4”と発声することができ、撮像デバイス260は4枚の写真を高速連写することができる。このコマンドは、ユーザが、シャッターボタン130を押下するよりも素速く複数の画像を撮影することを都合よく実現する。

【0065】

上述の通り、移動端末100は音声認識を用いて、撮像デバイス260に関する様々な制御動作を実行することができる。一部の実施形態において、移動端末100は、撮像デバイス260及び撮像デバイス260に付随する他の構成要素が音声コマンドによって制御されることを可能にするモード指定手段を含むことができる。例えば、移動端末100は、従前の手動起動モードでの使用と対となるものとして、撮像デバイス260の音声駆動制御モードでの使用を移動端末100のユーザが指定することを可能にするメニューを含むことができる。ユーザは、例えばユーザ制御エリア150(図1B)のボタンを押下して、そのメニューを受け取ることができる。あるいは、入力デバイス240が、撮像デバイス260を音声コマンドに関するモードと、従前のカメラ/ビデオ記録モードとのいずれとするかをユーザが選択することを可能にする、ユーザ制御エリア150に配置されるカメラ/ビデオボタンを含んでもよい。さらに別の代替実施形態においては、撮像デバ

10

20

30

40

50

イス260に関して音声駆動モードを設定する必要が無くてもよい。つまり、撮像デバイス260は、何らかの特定のモードを設定しなくても音声駆動モードで動作することができる。例えば、一実施形態において、移動端末100のユーザは、入力デバイス240に含まれていてよいマイクを起動するために、単に「マイクオン」と発声すればよい。マイクが起動すると、音声認識プログラム310はユーザからの音声コマンドを受信することができる。いずれの場合も、移動端末100は、撮像デバイス260に関する1つ又は複数の動作につながる音声コマンドをユーザが与えることを可能にする。

【0066】

結び

本発明に係る実施形態は、音声認識を用いることにより、撮像機能に関する制御動作を効率的な方法でユーザが実行することを可能にする。特に、音声認識を利用することで、カメラ及び/又はビデオ録画デバイスに関する様々な制御動作を、簡単でユーザフレンドリな方法によってユーザが実行することを可能にする。これにより、様々な撮影関連機能の利用についてのユーザ満足度を向上させることができ、それら機能の全般的な利用を簡単にすることができる。

10

【0067】

本発明の実施形態の記述は、例証及び説明を与えるものであるが、それが本発明の全てを余すところ無く開示しているという意図もなければ、本発明を開示されたそのものに限定するという意図もない。上述の教示を踏まえて変更物又は派生物を実現することも可能であるし、本発明の実践からも変更物又は派生物を得ることができる。

20

【0068】

例えば、本発明を主に、カメラ及び/又はビデオレコーダのような撮像デバイスを含んだ移動端末に関して説明してきた。しかし、本発明は、独立したカメラ、独立したビデオ録画デバイス、カメラ/ビデオ録画デバイスの組み合わせなどといった撮像機器を含んだ他の機器においても本発明を利用可能である。

【0069】

さらに、一部の実施形態において、移動端末100は、音声コマンドが特定された後で、移動端末100のユーザに対して音声及び/又はテキストのようなフィードバックを提供してもよい。例えば、ビデオモードへ切り替えるためのコマンドをユーザが発声した後、移動端末100は「ビデオモードが起動されました」と述べる音声を出力してもよい。このようにして、ユーザは、実行された特定の制御動作を示すフィードバックを受け取ることができる。

30

【0070】

図4及び図6に関し、一連の動作を説明したが、本発明に係る他の実施形態に実施形態において、それら動作の順序は変更されうる。さらに、依存関係にない動作は並行に実行されてもよい。

【0071】

上述した本発明の複数の見地が、セルラ通信機器/システム、方法、及び/又はコンピュータプログラム製品において実施可能であることは、本技術分野に属する当業者にとって自明であろう。従って、本発明はハードウェア及び/又はソフトウェア(ファームウェア、常駐ソフトウェア、マイクロコード等を含む)で実施可能である。さらに、本発明に係る見地は、コンピュータが利用可能な、又はコンピュータが読み取り可能な記憶媒体であって、コンピュータが利用可能な、又はコンピュータが読み取り可能なプログラムコードが当該記憶媒体内に実現されている記録媒体の状態のコンピュータプログラム製品の形態を取りうる。本発明は、本発明の原理に準拠した見地の実施に用いられる実際のソフトウェアコード又は専用の制御ハードウェアによって限定されない。従って、それら見地の動作及び振る舞いが、特定のソフトウェアコードに言及することなく説明されていれば、それは、本技術分野に属する当業者が、本明細書の記載に基づいてそれら見地を実施するためのソフトウェア及び制御ハードウェアを設計可能であるものと理解されるべきである。

40

50

【0072】

さらに、本発明の所定の部分は、1つ又は複数の機能を実行する「ロジック」として実現されうる。このロジックは、プロセッサ、マイクロプロセッサ、特定用途向け集積回路又はフィールドプログラマブルゲートアレイ、ソフトウェア、又はハードウェアとソフトウェアの組み合わせを含むことができる。

【0073】

なお、本明細書において「含む/有する」という語が用いられる場合、説明された機能、整数、ステップ又は構成部品の存在を特定するものとして解釈されるが、他の機能、整数、ステップ、構成部品の1つ又は複数、又はそれらグループ存在や付加を排除するものではないことを強調しておく。

【0074】

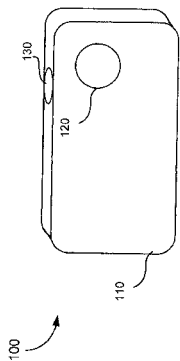
本願の説明に用いられているいかなる要素、動作、又は命令も、それが重要又は必須であると明記されていない限り、本発明にとって重要又は必須なものと解釈されるべきではない。また、複数と明記されていないものは、1つ又は複数の場合を含むことが意図されている。1つであることが意図されている場合には、「1つの」または類似の記載が用いられている。さらに、「~に基づく」という表現は、異なる意味が明記されていない限り、「少なくとも一部が~に基づく」という意味が意図されている。

【0075】

本発明の範囲は、添付の請求項とその等価物によって規定される。

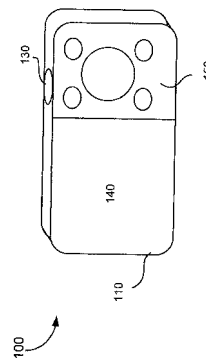
【図1A】

FIG. 1A



【図1B】

FIG. 1B



【図2】

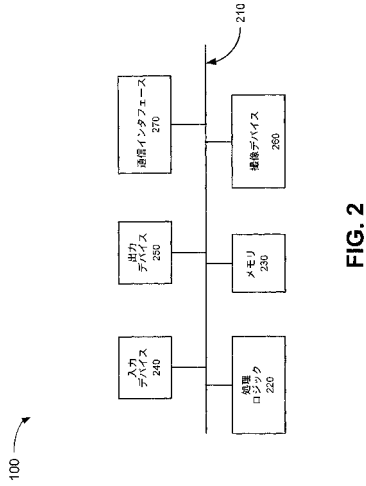


FIG. 2

【図3】

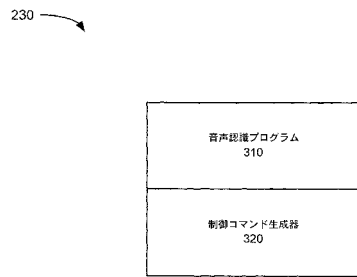


FIG. 3

【図5】

動作 510	音声コマンド 520
カメラのオン/オフ	カメラ オン/オフ
マイクのオン/オフ	マイク オン/オフ
撮影	チーズ
撮影時フラッシュ使用	フラッシュ
"N"枚の写真を高速連写	バースト "N"
所定量のズームイン/アウト	ズーム イン/アウト
最大のズームアウト	広角
通常のオートフォーカスオン/オフ	フォーカス オン/オフ
オートフォーカスマクロモードオン	オートフォーカス マクロ
ビデオ撮影開始/停止	ビデオ 開始/停止
最大解像度の使用	最大解像度

FIG. 5

【図4】

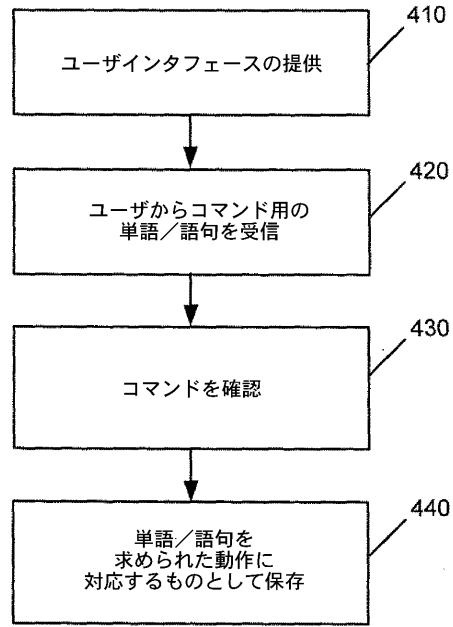


FIG. 4

【図6】

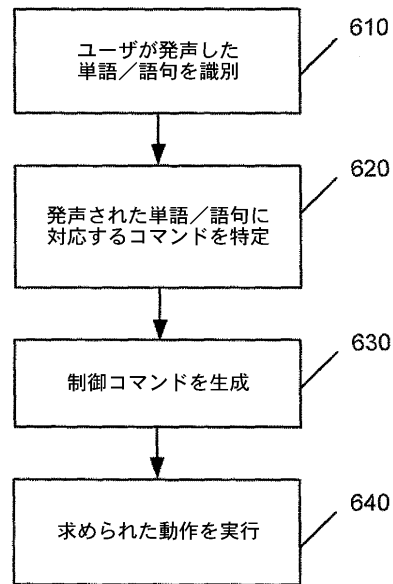


FIG. 6

フロントページの続き

(51)Int.Cl. F I
G 0 6 F 3/16 3 2 0 B

(72)発明者 ホープ, ジュリアン
イギリス国 ハーウッド B L 2 4 D J, ライディング ゲート, トップ オース ブロウ
コテージズ 17

審査官 田村 誠治

(56)参考文献 特開2005-027002(JP,A)
特開2001-216059(JP,A)
特開2000-083186(JP,A)
特開2004-180055(JP,A)
特開2002-312386(JP,A)

(58)調査した分野(Int.Cl., DB名)

H04N 5/225
G06F 3/16
G10L 15/00
G10L 15/06
H04N 5/232