

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6221535号
(P6221535)

(45) 発行日 平成29年11月1日(2017.11.1)

(24) 登録日 平成29年10月13日(2017.10.13)

(51) Int.Cl.

F I

G 1 O L 15/28 (2013.01)
 G 1 O L 21/0308 (2013.01)
 G O 6 F 3/0346 (2013.01)
 G O 6 F 3/0481 (2013.01)
 G O 6 F 3/16 (2006.01)

G 1 O L 15/28 2 3 O K
 G 1 O L 21/0308 Z
 G O 6 F 3/0346 4 2 3
 G O 6 F 3/0481 1 7 O
 G O 6 F 3/16 6 1 O

請求項の数 17 (全 31 頁) 最終頁に続く

(21) 出願番号 特願2013-188220 (P2013-188220)
 (22) 出願日 平成25年9月11日(2013.9.11)
 (65) 公開番号 特開2015-55718 (P2015-55718A)
 (43) 公開日 平成27年3月23日(2015.3.23)
 審査請求日 平成27年12月15日(2015.12.15)

(73) 特許権者 000002185
 ソニー株式会社
 東京都港区港南1丁目7番1号
 (74) 代理人 100095957
 弁理士 亀谷 美明
 (74) 代理人 100096389
 弁理士 金本 哲男
 (74) 代理人 100101557
 弁理士 萩原 康司
 (74) 代理人 100128587
 弁理士 松本 一騎
 (72) 発明者 井元 麻紀
 東京都港区港南1丁目7番1号 ソニー株
 式会社内

最終頁に続く

(54) 【発明の名称】 情報処理装置、情報処理方法、およびプログラム

(57) 【特許請求の範囲】

【請求項1】

表示画面におけるユーザの視線の位置に関する情報に基づいて、ユーザが所定のオブジェクトを見たかを判定する判定部と、

ユーザが所定のオブジェクトを見たときと判定された場合に、音声認識処理を制御する音声認識制御部と、

を備え、

前記音声認識制御部は、見たときと判定された前記所定のオブジェクトに基づいて、見たときと判定された前記所定のオブジェクトに対応する命令を特定し、特定された命令が認識されるように前記音声認識処理を制御し、

前記音声認識制御部は、

見たときと判定された前記所定のオブジェクトを含む表示画面における所定の領域内に含まれる、他のオブジェクトを判定し、

見たときと判定された前記所定のオブジェクトに対応付けられている命令と、判定された前記他のオブジェクトに対応付けられている命令とを、見たときと判定された前記所定のオブジェクトに対応する命令として特定する、情報処理装置。

【請求項2】

前記音声認識制御部は、

所定のオブジェクトを見たときと判定されたユーザに対応するユーザの視線の位置に関する情報に基づいて、音源分離を行うことが可能な音声入力デバイスに、所定のオブジェクト

を見たときと判定されたユーザの位置から発せられる音声を示す音声信号を取得させ、

前記音声入力デバイスにより取得された音声信号に対して音声認識を行わせる、請求項 1 に記載の情報処理装置。

【請求項 3】

前記音声認識制御部は、

所定のオブジェクトを見たときと判定されたユーザに対応するユーザの視線の位置に関する情報に基づくユーザの位置と、音源定位を行うことが可能な音声入力デバイスが測定した音源の位置との差分が、設定された閾値以下の場合、または、

前記ユーザの位置と前記音源の位置との差分が、前記閾値より小さい場合に、

前記音声入力デバイスにより取得された音声を示す音声信号に対して音声認識を行わせる、請求項 1 または 2 に記載の情報処理装置。

10

【請求項 4】

前記判定部は、ユーザの視線の位置に関する情報が示す視線の位置が、所定のオブジェクトを含む表示画面における第 1 領域内に含まれる場合に、ユーザが所定のオブジェクトを見たときと判定する、請求項 1 ～ 3 のいずれか 1 項に記載の情報処理装置。

【請求項 5】

前記判定部が、ユーザが所定のオブジェクトを見たときと判定した場合、

前記判定部は、所定のオブジェクトを見たときと判定されたユーザに対応するユーザの視線の位置に関する情報が示す視線の位置が、所定のオブジェクトを含む表示画面における第 2 領域内に含まれなくなったときに、前記ユーザが所定のオブジェクトを見ていないと判定し、

20

前記音声認識制御部は、前記ユーザが所定のオブジェクトを見ていないと判定されたときに、前記ユーザに対する音声認識を終了させる、請求項 1 ～ 4 のいずれか 1 項に記載の情報処理装置。

【請求項 6】

前記判定部が、ユーザが所定のオブジェクトを見たときと判定した場合、

前記判定部は、

所定のオブジェクトを見たときと判定されたユーザに対応するユーザの視線の位置に関する情報が示す視線の位置が、所定のオブジェクトを含む表示画面における第 2 領域内に含まれない状態が、設定された設定時間以上継続するとき、または、

30

所定のオブジェクトを見たときと判定されたユーザに対応するユーザの視線の位置に関する情報が示す視線の位置が前記第 2 領域内に含まれない状態が、前記設定時間より長く継続するとき、

前記ユーザが所定のオブジェクトを見ていないと判定し、

前記音声認識制御部は、前記ユーザが所定のオブジェクトを見ていないと判定されたときに、前記ユーザに対する音声認識を終了させる、請求項 1 ～ 4 のいずれか 1 項に記載の情報処理装置。

【請求項 7】

前記判定部は、所定のオブジェクトを見たときと判定されたユーザに対応するユーザの視線の位置に関する情報が示す視線の位置の履歴に基づいて、前記設定時間を動的に設定する、請求項 6 に記載の情報処理装置。

40

【請求項 8】

前記判定部は、一のユーザが所定のオブジェクトを見たときと判定した後に、前記一のユーザが所定のオブジェクトを見ていないと判定されていない場合には、他のユーザが所定のオブジェクトを見たときは判定しない、請求項 1 ～ 7 のいずれか 1 項に記載の情報処理装置。

【請求項 9】

前記判定部は、

前記表示画面において画像が表示される方向が撮像された撮像画像に基づいてユーザを特定し、

50

特定されたユーザに対応するユーザの視線の位置に関する情報に基づいて、ユーザが所定のオブジェクトを見たかを判定する、請求項 1 ~ 8 のいずれか 1 項に記載の情報処理装置。

【請求項 10】

前記所定のオブジェクトを表示画面に表示させる表示制御部をさらに備える、請求項 1 ~ 9 のいずれか 1 項に記載の情報処理装置。

【請求項 11】

前記表示制御部は、前記ユーザの視線の位置に関する情報が示す視線の位置によらず、表示画面における設定されている位置に、前記所定のオブジェクトを表示させる、請求項 10に記載の情報処理装置。

10

【請求項 12】

前記表示制御部は、前記ユーザの視線の位置に関する情報に基づいて、前記所定のオブジェクトを選択的に表示させる、請求項 10に記載の情報処理装置。

【請求項 13】

前記表示制御部は、前記所定のオブジェクトを表示させる場合には、設定されている表示方法を用いて前記所定のオブジェクトを表示させる、請求項 12に記載の情報処理装置。

【請求項 14】

前記表示制御部は、前記所定のオブジェクトを表示させる場合には、前記ユーザの視線の位置に関する情報が示す視線の位置に基づいて、段階的に前記所定のオブジェクトを表示させる、請求項 12または13に記載の情報処理装置。

20

【請求項 15】

前記表示制御部は、音声認識が行われている場合、前記所定のオブジェクトの表示態様を変える、請求項 10 ~ 14のいずれか 1 項に記載の情報処理装置。

【請求項 16】

表示画面におけるユーザの視線の位置に関する情報に基づいて、ユーザが所定のオブジェクトを見たかを判定するステップと、

ユーザが所定のオブジェクトを見たかと判定された場合に、音声認識処理を制御するステップと、

を有し、

30

前記音声認識処理を制御するステップでは、見たと判定された前記所定のオブジェクトに基づいて、見たと判定された前記所定のオブジェクトに対応する命令が特定され、特定された命令が認識されるように前記音声認識処理が制御され、

前記音声認識処理を制御するステップでは、

見たと判定された前記所定のオブジェクトを含む表示画面における所定の領域内に含まれる、他のオブジェクトが判定され、

見たと判定された前記所定のオブジェクトに対応付けられている命令と、判定された前記他のオブジェクトに対応付けられている命令とが、見たと判定された前記所定のオブジェクトに対応する命令として特定される、情報処理装置により実行される情報処理方法。

【請求項 17】

40

表示画面におけるユーザの視線の位置に関する情報に基づいて、ユーザが所定のオブジェクトを見たかを判定するステップ、

ユーザが所定のオブジェクトを見たかと判定された場合に、音声認識処理を制御するステップ、

をコンピュータに実行させ、

前記音声認識処理を制御するステップでは、見たと判定された前記所定のオブジェクトに基づいて、見たと判定された前記所定のオブジェクトに対応する命令が特定され、特定された命令が認識されるように前記音声認識処理が制御され、

前記音声認識処理を制御するステップでは、

見たと判定された前記所定のオブジェクトを含む表示画面における所定の領域内に含ま

50

れる、他のオブジェクトが判定され、

見たと判定された前記所定のオブジェクトに対応付けられている命令と、判定された前記他のオブジェクトに対応付けられている命令とが、見たと判定された前記所定のオブジェクトに対応する命令として特定される、プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本開示は、情報処理装置、情報処理方法、およびプログラムに関する。

【背景技術】

【0002】

近年、アイトラッキング技術などの視線検出技術を利用し、ユーザが視線により操作を行うことが可能なユーザインタフェースが登場している。ユーザが視線により操作を行うことが可能なユーザインタフェースに関する技術としては、例えば下記の特許文献1に記載の技術が挙げられる。

【先行技術文献】

【特許文献】

【0003】

【特許文献1】特開2009-64395号公報

【発明の概要】

【発明が解決しようとする課題】

【0004】

音声認識が行われる場合、音声認識を開始するトリガーとしては、例えば、ボタンを押すなどユーザが特定のユーザ操作を行うことや、ユーザが特定のワードを発話することが挙げられる。しかしながら、上記のような特定のユーザ操作や特定のワードの発話によって音声認識が行われる場合には、ユーザが行っていた操作や会話などを妨げる可能性がある。よって、上記のような特定のユーザ操作や特定のワードの発話によって音声認識が行われる場合には、ユーザの利便性を低下させる恐れがある。

【0005】

本開示では、音声認識が行われる場合におけるユーザの利便性の向上を図ることが可能な、新規かつ改良された情報処理装置、情報処理方法、およびプログラムを提案する。

【課題を解決するための手段】

【0006】

本開示によれば、表示画面におけるユーザの視線の位置に関する情報に基づいて、ユーザが所定のオブジェクトを見たかを判定する判定部と、ユーザが所定のオブジェクトを見たと判定された場合に、音声認識処理を制御する音声認識制御部と、を備える、情報処理装置が提供される。

【0007】

また、本開示によれば、表示画面におけるユーザの視線の位置に関する情報に基づいて、ユーザが所定のオブジェクトを見たかを判定するステップと、ユーザが所定のオブジェクトを見たと判定された場合に、音声認識処理を制御するステップと、を有する、情報処理装置により実行される情報処理方法が提供される。

【0008】

また、本開示によれば、表示画面におけるユーザの視線の位置に関する情報に基づいて、ユーザが所定のオブジェクトを見たかを判定するステップ、ユーザが所定のオブジェクトを見たと判定された場合に、音声認識処理を制御するステップ、をコンピュータに実行させるためのプログラムが提供される。

【発明の効果】

【0009】

本開示によれば、音声認識が行われる場合におけるユーザの利便性の向上を図ることができる。

10

20

30

40

50

【 0 0 1 0 】

なお、上記の効果は必ずしも限定的なものではなく、上記の効果とともに、または上記の効果に代えて、本明細書に示されたいずれかの効果、または本明細書から把握されうる他の効果が奏されてもよい。

【図面の簡単な説明】

【 0 0 1 1 】

【図 1】本実施形態に係る所定のオブジェクトの一例を示す説明図である。

【図 2】本実施形態に係る情報処理方法に係る処理の一例を説明するための説明図である。

【図 3】本実施形態に係る情報処理方法に係る処理の一例を説明するための説明図である 10

【図 4】本実施形態に係る情報処理方法に係る処理の一例を説明するための説明図である。

【図 5】本実施形態に係る情報処理方法に係る処理の一例を説明するための説明図である。

【図 6】本実施形態に係る情報処理方法に係る処理の一例を説明するための説明図である。

【図 7】本実施形態に係る情報処理方法に係る処理の一例を説明するための説明図である。

【図 8】本実施形態に係る情報処理装置の構成の一例を示すブロック図である。 20

【図 9】本実施形態に係る情報処理装置のハードウェア構成の一例を示す説明図である。

【発明を実施するための形態】

【 0 0 1 2 】

以下に添付図面を参照しながら、本開示の好適な実施の形態について詳細に説明する。なお、本明細書及び図面において、実質的に同一の機能構成を有する構成要素については、同一の符号を付することにより重複説明を省略する。

【 0 0 1 3 】

また、以下では、下記に示す順序で説明を行う。

1. 本実施形態に係る情報処理方法
2. 本実施形態に係る情報処理装置
3. 本実施形態に係るプログラム

30

【 0 0 1 4 】

(本実施形態に係る情報処理方法)

本実施形態に係る情報処理装置の構成について説明する前に、まず、本実施形態に係る情報処理方法について説明する。以下では、本実施形態に係る情報処理方法に係る処理を、本実施形態に係る情報処理装置が行う場合を例に挙げて、本実施形態に係る情報処理方法について説明する。

【 0 0 1 5 】

[1] 本実施形態に係る情報処理方法に係る処理の概要

上述したように、特定のユーザ操作や特定のワードの発話によって音声認識が行われる場合には、ユーザの利便性を低下させる恐れがある。また、特定のユーザ操作や特定のワードの発話を音声認識を開始するトリガーとする場合には、ユーザが行っていた他の操作や会話などを妨げる可能性があることから、特定のユーザ操作や特定のワードの発話は、自然な操作であるとは言い難い。

40

【 0 0 1 6 】

そこで、本実施形態に係る情報処理装置は、音声認識処理を制御することによって、特定のユーザ操作や特定のワードの発話が発検された場合に音声認識を行わせるのではなく、ユーザが表示画面に表示されている所定のオブジェクトを見たと判定された場合に、音声認識を行わせる。

【 0 0 1 7 】

50

ここで、本実施形態に係る情報処理装置が音声認識処理を制御する対象としては、例えば、自装置（本実施形態に係る情報処理装置。以下、同様とする。）や、通信部（後述する）や接続されている外部の通信デバイスを介して通信可能な外部装置が挙げられる。上記外部装置としては、例えば、サーバなど音声認識処理を行うことが可能な任意の装置が挙げられる。また、上記外部装置は、例えば、クラウドコンピューティングなどのように、ネットワークへの接続（または各装置間の通信）を前提とした、1または2以上の装置からなるシステムであってもよい。

【0018】

音声認識処理を制御する対象が自装置である場合、本実施形態に係る情報処理装置は、例えば、自装置において音声認識（音声認識処理）を行い、自装置において行われた音声認識の結果を利用する。本実施形態に係る情報処理装置は、例えば、音声認識することが可能な任意の技術を用いて音声を認識する。

10

【0019】

また、音声認識処理を制御する対象が上記外部装置である場合、本実施形態に係る情報処理装置は、例えば、音声認識を制御する命令を含む制御データを、通信部（後述する）などに、上記外部装置へと送信させる。本実施形態に係る音声認識を制御する命令としては、例えば、音声認識処理を行わせる命令と、音声認識処理を終了させる命令とが挙げられる。また、上記制御データには、例えば、さらにユーザが発話した音声を示す音声信号が含まれていてもよい。音声認識処理を行わせる命令を含む上記制御データを上記外部装置へと送信させる場合、本実施形態に係る情報処理装置は、例えば、上記外部装置から取得された“外部装置において行われた音声認識の結果を示すデータ”を利用する。

20

【0020】

以下では、本実施形態に係る情報処理装置が音声認識処理を制御する対象が、自装置である場合、すなわち、本実施形態に係る情報処理装置が音声認識を行う場合を主に例に挙げて、本実施形態に係る情報処理方法に係る処理について説明する。

【0021】

また、本実施形態に係る表示画面とは、例えば、様々な画像が表示され、ユーザが視線を向ける表示画面である。本実施形態に係る表示画面としては、例えば、本実施形態に係る情報処理装置が備える表示部（後述する）の表示画面や、本実施形態に係る情報処理装置と無線または有線で接続されている外部の表示装置（または、外部の表示デバイス）の表示画面などが挙げられる。

30

【0022】

図1は、本実施形態に係る所定のオブジェクトの一例を示す説明図である。図1のA～図1のCは、表示画面に表示されている、所定のオブジェクトを含む画像の一例をそれぞれ示している。

【0023】

本実施形態に係る所定のオブジェクトとしては、例えば、図1のAのO1に示すような音声認識を行わせるためのアイコン（以下、「音声認識アイコン」と示す。）や、図1のBのO2に示すような音声認識を行わせるための画像（以下、「音声認識画像」と示す。）が挙げられる。図1のBに示す例では、本実施形態に係る音声認識画像として、キャラクターを示すキャラクター画像を示している。なお、本実施形態に係る音声認識アイコンや、本実施形態に係る音声認識画像が、図1のAや図1のBに示す例に限られないことは言うまでもない。

40

【0024】

なお、本実施形態に係る所定のオブジェクトは、音声認識アイコンや音声認識画像に限られない。例えば、本実施形態に係る所定のオブジェクトは、例えば図1のCのO3に示すオブジェクトのように、ユーザ操作により選択されうるオブジェクト（以下、「選択候補オブジェクト」と示す。）であってもよい。図1のCに示す例では、本実施形態に係る選択候補オブジェクトとして、映画のタイトルなどを示すサムネイル画像を示している。なお、図1のCでは、符号O3が付されていないサムネイル画像やアイコンが、本実施形

50

態に係る選択候補オブジェクトであってもよい。また、本実施形態に係る選択候補オブジェクトが、図1のCに示す例に限られないことは言うまでもない。

【0025】

本実施形態に係る情報処理装置が、ユーザが表示画面に表示されている図1に示すような所定のオブジェクトを見たときと判定したときに、音声認識を行う場合、ユーザは、例えば、所定のオブジェクトに視線を向けて所定のオブジェクトを見ることによって、本実施形態に係る情報処理装置に音声認識を開始させることが可能となる。

【0026】

また、仮に、ユーザが、他の操作や会話などを行っていたとしても、ユーザが所定のオブジェクトを見ることにより当該他の操作や会話を妨げる可能性は、特定のユーザ操作や特定のワードの発話によって音声認識が行われる場合よりも低い。

10

【0027】

さらに、ユーザが表示画面に表示されている所定のオブジェクトを見たことを、音声認識を開始するトリガーとする場合には、ユーザが行っていた他の操作や会話などを妨げる可能性が低いことから、ユーザが表示画面に表示されている所定のオブジェクトを見ることは、上記特定のユーザ操作や特定のワードの発話よりも、より自然な操作であるといえる。

【0028】

したがって、本実施形態に係る情報処理装置が、本実施形態に係る情報処理方法に係る処理として、ユーザが表示画面に表示されている所定のオブジェクトを見たときと判定したときに音声認識を行わせることによって、音声認識が行われる場合におけるユーザの利便性の向上を図ることができる。

20

【0029】

[2] 本実施形態に係る情報処理方法に係る処理

次に、本実施形態に係る情報処理方法に係る処理について、より具体的に説明する。

【0030】

本実施形態に係る情報処理装置は、例えば、本実施形態に係る情報処理方法に係る処理として、例えば、下記に示す(1)判定処理、および(2)音声認識制御処理を行うことによって、ユーザの利便性の向上を図る。

【0031】

30

(1) 判定処理

本実施形態に係る情報処理装置は、例えば、表示画面におけるユーザの視線の位置に関する情報に基づいて、ユーザが所定のオブジェクトを見たかを判定する。

【0032】

ここで、本実施形態に係るユーザの視線の位置に関する情報とは、例えば、ユーザの視線の位置を示すデータ、または、ユーザの視線の位置の特定に用いることが可能なデータ(または、ユーザの視線の位置の推定に用いることが可能なデータ。以下、同様とする。)である。

【0033】

本実施形態に係るユーザの視線の位置を示すデータとしては、例えば、表示画面におけるユーザの視線の位置を示す座標データが挙げられる。表示画面におけるユーザの視線の位置は、例えば、表示画面における基準位置を原点とする座標系における座標で表される。また、本実施形態に係るユーザの視線の位置を示すデータには、視線の方向を示すデータ(例えば、表示画面に対する角度を示すデータなど)が含まれていてもよい。

40

【0034】

また、本実施形態に係るユーザの視線の位置の特定に用いることが可能なデータとしては、例えば、表示画面において画像(動画像または静止画像)が表示される方向が撮像された撮像画像データが挙げられる。また、本実施形態に係るユーザの視線の位置の特定に用いることが可能なデータには、さらに、表示画面において画像が表示される方向の赤外線を検出する赤外線センサの検出データなど、ユーザの視線の位置の推定精度の向上に用

50

いることが可能な検出値を得る、任意のセンサの検出データが含まれていてもよい。

【0035】

本実施形態に係るユーザの視線の位置に関する情報として、表示画面におけるユーザの視線の位置を示す座標データが用いられる場合、本実施形態に係る情報処理装置は、例えば、視線検出技術を用いてユーザの視線の位置を特定（または推定）した外部装置から取得された、表示画面におけるユーザの視線の位置を示す座標データを用いて、表示画面におけるユーザの視線の位置を特定する。また、本実施形態に係るユーザの視線の位置に関する情報として、視線の方向を示すデータが用いられる場合、本実施形態に係る情報処理装置は、例えば、上記外部装置から取得された視線の方向を示すデータを用いて、ユーザの視線の方向を特定する。

10

【0036】

ここで、例えば、視線検出技術を用いて検出された視線と、表示画面において画像が表示される方向が撮像された撮像画像から検出される表示画面に対するユーザの位置や顔の向きなどを用いることによって、表示画面におけるユーザの視線の位置や、ユーザの視線の方向を特定することが可能である。なお、本実施形態に係る表示画面におけるユーザの視線の位置やユーザの視線の方向の特定方法は、上記に限られない。例えば、本実施形態に係る情報処理装置や、外部装置は、表示画面におけるユーザの視線の位置やユーザの視線の方向を特定することが可能な任意の技術を用いることが可能である。

【0037】

また、本実施形態に係る視線検出技術としては、例えば、目の基準点（例えば、目頭や角膜反射などの目における動かない部分に対応する点）に対する、目の動点（例えば、虹彩や瞳孔などの目における動く部分に対応する点）の位置に基づいて、視線を検出する方法が挙げられる。なお、本実施形態に係る視線検出技術は、上記に限られず、例えば、視線を検出することが可能な任意の視線検出技術であってもよい。

20

【0038】

本実施形態に係るユーザの視線の位置に関する情報として、ユーザの視線の位置の特定に用いることが可能なデータが用いられる場合、本実施形態に係る情報処理装置は、例えば、自装置が備える撮像部（後述する）や外部の撮像デバイスから取得された撮像画像データ（ユーザの視線の位置の特定に用いることが可能なデータの一例）を用いる。また、上記の場合、本実施形態に係る情報処理装置は、例えば、自装置が備えるユーザの視線の位置の推定精度の向上に用いることが可能なセンサや、外部のセンサから取得された検出データ（ユーザの視線の位置の特定に用いることが可能なデータの一例）を用いてもよい。本実施形態に係る情報処理装置は、例えば上記のように取得されたユーザの視線の位置の特定に用いることが可能なデータを用いて、本実施形態に係る表示画面におけるユーザの視線の位置やユーザの視線の方向の特定方法に係る処理を行い、表示画面におけるユーザの視線の位置やユーザの視線の方向を特定する。

30

【0039】

（1-1）判定処理の第1の例

本実施形態に係る情報処理装置は、例えば、ユーザの視線の位置に関する情報が示す視線の位置が、所定のオブジェクトを含む表示画面における第1領域内に含まれる場合に、ユーザが所定のオブジェクトを見たと判定する。

40

【0040】

ここで、本実施形態に係る第1領域は、例えば、所定のオブジェクトにおける基準位置を基に設定される。本実施形態に係る基準位置としては、例えば、オブジェクトの中心点など、予め設定されているオブジェクト内の任意の位置が挙げられる。本実施形態に係る第1領域の大きさや形状は、予め設定されていてもよいし、ユーザ操作などに基づいて設定されてもよい。一例を挙げると、本実施形態に係る第1領域としては、例えば、所定のオブジェクトを含む領域のうちの最小の領域（すなわち、所定のオブジェクトが表示される領域）や、所定のオブジェクトの基準点を中心とする円形や矩形の領域などが挙げられる。また、本実施形態に係る第1領域は、例えば、表示画面における表示領域が分割され

50

た領域（以下、「分割領域」と示す。）であってもよい。

【0041】

より具体的には、本実施形態に係る情報処理装置は、例えば、ユーザの視線の位置に関する情報が示す視線の位置が、所定のオブジェクトを含む表示画面における第1領域内に含まれたときに、ユーザが所定のオブジェクトを見たかと判定する。

【0042】

なお、第1の例に係る判定処理は、上記に限られない。

【0043】

例えば、本実施形態に係る情報処理装置は、例えば、ユーザの視線の位置に関する情報が示す視線の位置が、第1領域内である時間が、設定されている第1設定時間よりも長い場合に、ユーザが所定のオブジェクトを見たかと判定してもよい。また、本実施形態に係る情報処理装置は、例えば、ユーザの視線の位置に関する情報が示す視線の位置が第1領域内である時間が、第1設定時間以上である場合に、ユーザが所定のオブジェクトを見たかと判定することも可能である。

10

【0044】

本実施形態に係る第1設定時間としては、例えば、本実施形態に係る情報処理装置の製造者やユーザ操作などに基づいて予め設定されている時間が挙げられる。本実施形態に係る第1設定時間が、予め設定されている時間である場合、本実施形態に係る情報処理装置は、ユーザの視線の位置に関する情報が示す視線の位置が第1領域内である時間と、予め設定されている第1設定時間とに基づいて、ユーザが所定のオブジェクトを見たかを判定する。

20

【0045】

本実施形態に係る情報処理装置は、例えば上記第1の例に係る判定処理を行うことによって、ユーザの視線の位置に関する情報に基づいて、ユーザが所定のオブジェクトを見たか否かを判定する。

【0046】

上述したように、本実施形態に係る情報処理装置は、ユーザが表示画面に表示されている所定のオブジェクトを見たかと判定された場合に、音声認識を行わせる。つまり、本実施形態に係る情報処理装置は、例えば上記第1の例に係る判定処理を行った結果、ユーザが所定のオブジェクトを見たかと判定された場合に、後述する(2)の処理(音声認識制御処理)を開始して音声認識を行わせる。

30

【0047】

なお、本実施形態に係る判定処理は、上記第1の例に係る判定処理のように、ユーザが所定のオブジェクトを見たかを判定する処理に限られない。

【0048】

例えば、本実施形態に係る情報処理装置は、ユーザの視線の位置に関する情報に基づいて、ユーザが所定のオブジェクトを見たかと判定された後に、ユーザが所定のオブジェクトを見ていないことを判定する。第2の例に係る判定処理において、ユーザが所定のオブジェクトを見たかと判定された後に、当該ユーザが所定のオブジェクトを見ていないと判定された場合には、後述する(2)の処理(音声認識制御処理)では、当該ユーザに対する音声認識を終了させることとなる。

40

【0049】

具体的には、ユーザが所定のオブジェクトを見たかと判定した場合、本実施形態に係る情報処理装置は、例えば、下記の第2の例に係る判定処理や、下記の第3の例に係る判定処理を行うことによって、ユーザが所定のオブジェクトを見ていないことを判定する。

【0050】

(1-2)判定処理の第2の例

本実施形態に係る情報処理装置は、例えば、所定のオブジェクトを見たかと判定されたユーザに対応するユーザの視線の位置に関する情報が示す視線の位置が、所定のオブジェクトを含む表示画面における第2領域内に含まれなくなったときに、ユーザが所定のオブジ

50

ェクトを見ていないと判定する。

【 0 0 5 1 】

本実施形態に係る第 2 領域としては、例えば、本実施形態に係る第 1 領域と同一の領域が挙げられる。なお、本実施形態に係る第 2 領域は、上記に限られない。例えば、本実施形態に係る第 2 領域は、本実施形態に係る第 1 領域よりも大きな領域であってもよい。

【 0 0 5 2 】

一例を挙げると、本実施形態に係る第 2 領域としては、例えば、所定のオブジェクトを含む領域のうちの最小の領域（すなわち、所定のオブジェクトが表示される領域）や、所定のオブジェクトの基準点を中心とする円形や矩形の領域などが挙げられる。また、本実施形態に係る第 2 領域は、例えば、分割領域であってもよい。本実施形態に係る第 2 領域の具体例については、後述する。

10

【 0 0 5 3 】

例えば、本実施形態に係る第 1 領域と本実施形態に係る第 2 領域とが、共に所定のオブジェクトを含む領域のうちの最小の領域（すなわち、所定のオブジェクトが表示される領域）であるときには、本実施形態に係る情報処理装置は、ユーザが所定のオブジェクトから目を逸らした場合に、ユーザが所定のオブジェクトを見ていないと判定する。そして、本実施形態に係る情報処理装置は、後述する（ 2 ）の処理（音声認識制御処理）において、上記ユーザに対する音声認識を終了させる。

【 0 0 5 4 】

また、例えば、本実施形態に係る第 2 領域が、上記最小の領域より大きい領域であるときには、本実施形態に係る情報処理装置は、ユーザが第 2 領域から目を逸らした場合に、ユーザが所定のオブジェクトを見ていないと判定する。そして、本実施形態に係る情報処理装置は、後述する（ 2 ）の処理（音声認識制御処理）において、上記ユーザに対する音声認識を終了させる。

20

【 0 0 5 5 】

図 2 は、本実施形態に係る情報処理方法に係る処理の一例を説明するための説明図である。図 2 は、表示画面に表示された画像の一例を示している。また、図 2 では、本実施形態に係る所定のオブジェクトを符号 O で表しており、所定のオブジェクトが音声認識アイコンである例を示している。以下では、本実施形態に係る所定のオブジェクトを、「所定のオブジェクト O」と示す場合がある。図 2 に示す領域 R 1 ~ R 3 は、表示画面の表示領域を 3 分割した領域であり、本実施形態に係る分割領域に該当する。

30

【 0 0 5 6 】

例えば、本実施形態に係る第 2 領域が、分割領域 R 1 である場合、本実施形態に係る情報処理装置は、ユーザが分割領域 R 1 から目を逸らした場合に、ユーザが所定のオブジェクト O 1 を見ていないと判定する。そして、本実施形態に係る情報処理装置は、後述する（ 2 ）の処理（音声認識制御処理）において、上記ユーザに対する音声認識を終了させる。

【 0 0 5 7 】

本実施形態に係る情報処理装置は、例えば、図 2 に示す分割領域 R 1 のように、設定されている第 2 領域に基づいて、ユーザが所定のオブジェクト O 1 を見ていないことを判定する。なお、本実施形態に係る第 2 領域が、図 2 に示す例に限られないことは、言うまでもない。

40

【 0 0 5 8 】

（ 1 - 3 ）判定処理の第 3 の例

本実施形態に係る情報処理装置は、例えば、所定のオブジェクトを見たときと判定されたユーザに対応するユーザの視線の位置に関する情報が示す視線の位置が、所定の領域内に含まれない状態が、設定された第 2 設定時間以上継続する場合に、ユーザが所定のオブジェクトを見ていないと判定する。また、本実施形態に係る情報処理装置は、例えば所定のオブジェクトを見たときと判定されたユーザに対応するユーザの視線の位置に関する情報が示す視線の位置が、所定の領域内に含まれない状態が、第 2 設定時間より長く継続する場合に

50

、ユーザが所定のオブジェクトを見ていないと判定してもよい。

【 0 0 5 9 】

本実施形態に係る第2設定時間としては、例えば、本実施形態に係る情報処理装置の製造者やユーザ操作などに基づいて予め設定されている時間が挙げられる。本実施形態に係る第2設定時間が、予め設定されている時間である場合、本実施形態に係る情報処理装置は、ユーザの視線の位置に関する情報が示す視線の位置が第2領域に含まれなくなっからの時間と、予め設定されている第2設定時間とに基づいて、ユーザが所定のオブジェクトを見ていないことを判定する。

【 0 0 6 0 】

なお、本実施形態に係る第2設定時間は、予め設定されている時間に限られない。

10

【 0 0 6 1 】

例えば、本実施形態に係る情報処理装置は、所定のオブジェクトを見たと判定されたユーザに対応するユーザの視線の位置に関する情報が示す視線の位置の履歴に基づいて、第2設定時間を動的に設定することも可能である。

【 0 0 6 2 】

本実施形態に係る情報処理装置は、例えば、ユーザの視線の位置に関する情報を、記憶部（後述する）や外部記録媒体などの記録媒体に逐次記録する。また、本実施形態に係る情報処理装置は、例えば、上記記録媒体に記憶されてから設定されている所定の時間が経過したユーザの視線の位置に関する情報を、上記記録媒体から削除してもよい。

【 0 0 6 3 】

20

そして、本実施形態に係る情報処理装置は、上記記録媒体に逐次記憶されているユーザの視線の位置に関する情報（すなわち、ユーザの視線の位置の履歴を示す、ユーザの視線の位置に関する情報。以下、「履歴情報」と示す。）を用いて、第2設定時間を動的に設定する。

【 0 0 6 4 】

例えば、本実施形態に係る情報処理装置は、履歴情報の中に、履歴情報が示すユーザの視線の位置と第2領域の境界部分との距離が、設定されている所定の距離以下の履歴情報が存在する場合に、第2設定時間を長くする。また、本実施形態に係る情報処理装置は、例えば、履歴情報の中に、履歴情報が示すユーザの視線の位置と第2領域の境界部分との距離が、設定されている所定の距離より小さい履歴情報が存在する場合に、第2設定時間を長くしてもよい。

30

【 0 0 6 5 】

本実施形態に係る情報処理装置は、例えば、第2設定時間を設定されている固定の時間分長くする。また、本実施形態に係る情報処理装置は、例えば、上記距離以下の履歴情報（または、上記距離より小さい履歴情報）のデータ数に応じて第2設定時間を長くする時間を変えてもよい。

【 0 0 6 6 】

例えば上記のように、第2設定時間が動的に設定されることによって、本実施形態に係る情報処理装置は、ユーザが所定のオブジェクトを見ていないことを判定する場合においてヒステリシスを考慮することができる。

40

【 0 0 6 7 】

なお、本実施形態に係る判定処理は、上記第1の例に係る判定処理～上記第3の例に係る判定処理に限られない。

【 0 0 6 8 】

（ 1 - 4 ）判定処理の第4の例

本実施形態に係る情報処理装置は、例えば、一のユーザが所定のオブジェクトを見た判定した後に、当該一のユーザが所定のオブジェクトを見ていないと判定されていない場合には、他のユーザが所定のオブジェクトを見たとは判定しない。

【 0 0 6 9 】

例えば、後述する（ 2 ）の処理（音声認識制御処理）において音声認識を行わせるとき

50

、処理を行う音声による命令が、機器の操作に関する命令である場合には、一度に受け入れられる音声による命令が1つである方が、望ましい。一度に受け入れる音声による命令が複数である場合には、例えば、相反する命令が連続して実行されるなど、ユーザの利便性の低下を招く恐れがあるからである。

【0070】

本実施形態に係る情報処理装置が、第4の例に係る判定処理を行うことによって、仮に、他のユーザが所定のオブジェクトを見た場合であっても、当該他のユーザが所定のオブジェクトを見たとは判定されないの、上記のようなユーザの利便性の低下を招く恐れがある事態を防止することができる。

【0071】

10

(1-5) 判定処理の第5の例

また、本実施形態に係る情報処理装置は、ユーザを特定し、特定されたユーザに対応するユーザの視線の位置に関する情報に基づいて、ユーザが所定のオブジェクトを見たかを判定してもよい。

【0072】

本実施形態に係る情報処理装置は、例えば、表示画面において画像が表示される方向が撮像された撮像画像に基づいてユーザを特定する。具体的には、本実施形態に係る情報処理装置は、例えば、撮像画像に対して顔認識処理を行うことによって、ユーザを特定するが、ユーザの特定方法は、上記に限られない。

【0073】

20

ユーザが特定されると、本実施形態に係る情報処理装置は、例えば、特定したユーザに対応するユーザIDを認識し、認識したユーザIDに対応するユーザの視線の位置に関する情報に基づいて、上記第1の例に係る判定処理と同様の処理を行う。

【0074】

(2) 音声認識制御処理

本実施形態に係る情報処理装置は、例えば、上記(1)の処理(判定処理)においてユーザが所定のオブジェクトを見たとは判定された場合に、音声認識処理を制御し、音声認識を行わせる。

【0075】

より具体的には、本実施形態に係る情報処理装置は、例えば下記の第1の例に係る音声認識制御処理や第2の例に係る音声認識制御処理に示すように、音源分離や音源定位を利用して、音声認識を行わせる。ここで、本実施形態に係る音源分離とは、様々な音の中から目的の音声のみを抽出する技術をいう。また、本実施形態に係る音源定位とは、音源の位置(角度)を測定する技術をいう。

【0076】

30

(2-1) 音声認識制御処理の第1の例：音源分離が利用される場合

本実施形態に係る情報処理装置は、音源分離を行うことが可能な音声入力デバイスと連携して、音声認識を行わせる。本実施形態に係る音源分離を行うことが可能な音声入力デバイスは、例えば、本実施形態に係る情報処理装置が備える音声入力デバイスであってもよいし、本実施形態に係る情報処理装置の外部の音声入力デバイスであってもよい。

【0077】

40

本実施形態に係る情報処理装置は、例えば、所定のオブジェクトを見たとは判定されたユーザに対応するユーザの視線の位置に関する情報に基づいて、音源分離を行うことが可能な音声入力デバイスに、所定のオブジェクトを見たとは判定されたユーザの位置から発せられる音声を示す音声信号を取得させる。そして、本実施形態に係る情報処理装置は、上記音声入力デバイスにより取得された音声信号に対して音声認識を行わせる。

【0078】

本実施形態に係る情報処理装置は、例えば、所定のオブジェクトを見たとは判定されたユーザに対応するユーザの視線の位置に関する情報に基づいて、当該ユーザの視線の向き(例えば、表示画面に対する視線の角度)を算出する。また、ユーザの視線の位置に関する

50

情報に視線の方向を示すデータが含まれる場合には、本実施形態に係る情報処理装置は、当該視線の方向を示すデータが示すユーザの視線の向きを用いる。そして、本実施形態に係る情報処理装置は、例えば、算出することなどにより得られたユーザの視線の向きに対して音源分離を行わせるための制御命令を、音源分離を行うことが可能な音声入力デバイスに送信する。上記制御命令に基づき音源分離を行うことによって、上記音声入力デバイスは、所定のオブジェクトを見たと判定されたユーザの位置から発せられる音声を示す音声信号を取得する。なお、本実施形態に係る音源分離を行うことが可能な音声入力デバイスにおける音声信号の取得方法が、上記に限られないことは、言うまでもない。

【0079】

図3は、本実施形態に係る情報処理方法に係る処理の一例を説明するための説明図であり、音声認識制御処理において音源分離が利用される場合の概要を示している。図3に示すD1は、表示画面を表示させる表示デバイスの一例を示しており、図3に示すD2は、音源分離を行うことが可能な音声入力デバイスの一例を示している。また、図3では、所定のオブジェクトOが音声認識アイコンである例を示している。また、図3では、ユーザU1～U3という3人のユーザが、それぞれ表示画面を見ている例を示している。また、図3のCに示すR0は、音声入力デバイスD2が音声を取得することが可能な領域の一例を示しており、図3のCに示すR1は、音声入力デバイスD2が音声を取得する領域の一例を示している。図3では、本実施形態に係る情報処理方法に係る処理の流れを、図3に示すA、図3に示すB、図3に示すCの順番で時系列に表している。

【0080】

ユーザU1～U3それぞれが表示画面をみている場合において、例えば、ユーザU1が、表示画面の右端を見ると（図3に示すA）、本実施形態に係る情報処理装置は、表示画面に所定のオブジェクトOを表示する（図3に示すB）。本実施形態に係る情報処理装置は、例えば、後述する本実施形態に係る表示制御処理を行うことによって、表示画面に所定のオブジェクトOを表示する。

【0081】

表示画面に所定のオブジェクトOを表示すると、本実施形態に係る情報処理装置は、例えば、上記（1）の処理（判定処理）を行うことによって、ユーザが所定のオブジェクトOを見ているかを判定する。図3のBに示す例では、本実施形態に係る情報処理装置は、ユーザU1が所定のオブジェクトOを見たと判定する。

【0082】

ユーザU1が所定のオブジェクトOを見たと判定されると、本実施形態に係る情報処理装置は、ユーザU1に対応するユーザの視線の位置に関する情報に基づく制御命令を、音源分離を行うことが可能な音声入力デバイスD2に送信する。音声入力デバイスD2は、上記制御命令に基づいて、所定のオブジェクトを見たと判定されたユーザの位置から発せられる音声を示す音声信号を取得する（図3のC）。そして、本実施形態に係る情報処理装置は、音声入力デバイスD2から音声信号を取得する。

【0083】

音声入力デバイスD2から音声信号が取得されると、本実施形態に係る情報処理装置は、当該音声信号に対して音声認識に係る処理（後述する）を行い、音声認識に係る処理の結果、認識された命令を実行する。

【0084】

音源分離が利用される場合、本実施形態に係る情報処理装置は、本実施形態に係る情報処理方法に係る処理として、例えば図3を参照して示したような処理を行う。なお、音源分離が利用される場合における本実施形態に係る情報処理方法に係る処理の例が、図3を参照して示した例に限られないことは、言うまでもない。

【0085】

（2-2）音声認識制御処理の第2の例：音源定位が利用される場合

本実施形態に係る情報処理装置は、音源定位を行うことが可能な音声入力デバイスと連携して、音声認識を行わせる。本実施形態に係る音源定位を行うことが可能な音声入力デ

10

20

30

40

50

バイスは、例えば、本実施形態に係る情報処理装置が備える音声入力デバイスであってもよいし、本実施形態に係る情報処理装置の外部の音声入力デバイスであってもよい。

【0086】

本実施形態に係る情報処理装置は、例えば、所定のオブジェクトを見たと判定されたユーザに対応するユーザの視線の位置に関する情報に基づくユーザの位置と、音源定位を行うことが可能な音声入力デバイスが測定した音源の位置との差分に基づいて、音源定位を行うことが可能な音声入力デバイスにより取得された音声を示す音声信号に対して選択的に音声認識を行わせる。

【0087】

より具体的には、本実施形態に係る情報処理装置は、例えば、上記ユーザの視線の位置に関する情報に基づくユーザの位置と上記音源の位置との差分が、設定された閾値以下の場合（または、上記ユーザの視線の位置に関する情報に基づくユーザの位置と上記音源の位置との差分が、当該閾値より小さい場合。以下、同様とする。）に、上記音声信号に対して選択的に音声認識を行わせる。ここで、第2の例に係る音声認識制御処理に係る閾値は、例えば、予め設定されている固定値であってもよいし、ユーザ操作などに基づき変更可能な可変値であってもよい。

【0088】

本実施形態に係る情報処理装置は、例えば、音源定位を行うことが可能な音声入力デバイスから適宜送信される、音源の位置を示す情報（データ）を用いる。また、本実施形態に係る情報処理装置は、例えば、上記（1）の処理（判定処理）においてユーザが所定のオブジェクトを見ていると判定された場合に、音源定位を行うことが可能な音声入力デバイスに対して、音源の位置を示す情報の送信を要求する命令を送信し、当該命令に応じて音声入力デバイスから送信された音源の位置を示す情報を用いることも可能である。

【0089】

図4は、本実施形態に係る情報処理方法に係る処理の一例を説明するための説明図であり、音声認識制御処理において音源定位が利用される場合の概要を示している。図4に示すD1は、表示画面を表示させる表示デバイスの一例を示しており、図4に示すD2は、音源分離を行うことが可能な音声入力デバイスの一例を示している。また、図4では、所定のオブジェクトOが音声認識アイコンである例を示している。また、図4では、ユーザU1～U3という3人のユーザが、それぞれ表示画面を見ている例を示している。また、図4のCに示すR0は、音声入力デバイスD2が音源定位を行うことが可能な領域の一例を示しており、図4のCに示すR2は、音声入力デバイスD2により特定された音源の位置の一例を示している。図4では、本実施形態に係る情報処理方法に係る処理の流れを、図4に示すA、図4に示すB、図4に示すCの順番で時系列に表している。

【0090】

ユーザU1～U3それぞれが表示画面をみている場合において、例えば、ユーザU1が、表示画面の右端を見ると（図4に示すA）、本実施形態に係る情報処理装置は、表示画面に所定のオブジェクトOを表示する（図4に示すB）。本実施形態に係る情報処理装置は、例えば、後述する本実施形態に係る表示制御処理を行うことによって、表示画面に所定のオブジェクトOを表示する。

【0091】

表示画面に所定のオブジェクトOを表示すると、本実施形態に係る情報処理装置は、例えば、上記（1）の処理（判定処理）を行うことによって、ユーザが所定のオブジェクトOを見ているかを判定する。図4のBに示す例では、本実施形態に係る情報処理装置は、ユーザU1が所定のオブジェクトOを見たと判定する。

【0092】

ユーザU1が所定のオブジェクトOを見たと判定されると、本実施形態に係る情報処理装置は、所定のオブジェクトを見たと判定されたユーザに対応するユーザの視線の位置に関する情報に基づくユーザの位置と、音源定位を行うことが可能な音声入力デバイスが測定した音源の位置との差分を算出する。ここで、本実施形態に係るユーザの視線の位置に

10

20

30

40

50

関する情報に基づくユーザの位置と、音声入力デバイスが測定した音源の位置とは、例えば、表示画面に対する角度で表される。なお、本実施形態に係るユーザの視線の位置に関する情報に基づくユーザの位置と、音声入力デバイスが測定した音源の位置とは、表示画面に対応する平面を示す 2 軸と、表示画面に対する垂直方向を示す 1 軸とからなる 3 次元座標系の座標で表されていてもよい。

【 0 0 9 3 】

本実施形態に係る情報処理装置は、例えば、算出された差分が設定されている閾値以下の場合に、音源定位を行うことが可能な音声入力デバイス D 2 により取得された音声を示す音声信号に対して音声認識に係る処理（後述する）を行う。そして、本実施形態に係る情報処理装置は、音声認識に係る処理の結果、認識された命令を実行する。

10

【 0 0 9 4 】

音源定位が利用される場合、本実施形態に係る情報処理装置は、本実施形態に係る情報処理方法に係る処理として、例えば図 4 を参照して示したような処理を行う。なお、音源定位が利用される場合における本実施形態に係る情報処理方法に係る処理の例が、図 4 を参照して示した例に限られないことは、言うまでもない。

【 0 0 9 5 】

本実施形態に係る情報処理装置は、例えば、上記（ 2 - 1 ）に示す第 1 の例に係る音声認識制御処理や、上記（ 2 - 2 ）に示す第 2 の例に係る音声認識制御処理に示すように、音源分離や音源定位を利用して音声認識を行わせる。

【 0 0 9 6 】

20

次に、本実施形態に係る音声認識制御処理における、音声認識に係る処理について説明する。

【 0 0 9 7 】

本実施形態に係る情報処理装置は、取得された音声信号から、上記（ 1 ）の処理（判定処理）においてユーザが見たと判定された所定のオブジェクトによらずに、認識可能な全ての命令を認識する。そして、本実施形態に係る情報処理装置は、認識された命令を実行する。

【 0 0 9 8 】

なお、本実施形態に係る音声認識に係る処理において認識する命令は、上記に限られない。

30

【 0 0 9 9 】

例えば、本実施形態に係る情報処理装置は、上記（ 1 ）の処理（判定処理）においてユーザが見たと判定された所定のオブジェクトに基づいて、認識する命令を動的に変えるように制御することも可能である。本実施形態に係る情報処理装置は、例えば、上述した音声認識処理を制御する対象と同様に、自装置や、通信部（後述する）や接続されている外部の通信デバイスを介して通信可能な外部装置を、認識する命令を動的に変える制御の制御対象とする。より具体的には、本実施形態に係る情報処理装置は、例えば下記の（ A ） 、 （ B ）に示すように、認識する命令を動的に変えるように制御する。

【 0 1 0 0 】

（ A ）本実施形態に係る音声認識に係る処理における、認識する命令の動的な変更の第 1 の例

40

本実施形態に係る情報処理装置は、上記（ 1 ）の処理（判定処理）においてユーザが見たと判定された所定のオブジェクトに対応する命令を認識するように制御する。

【 0 1 0 1 】

（ A - 1 ）

認識する命令を動的に変える制御の制御対象が自装置である場合、本実施形態に係る情報処理装置は、例えば、オブジェクトと命令（または、命令群）とが対応付けられているテーブル（または、データベース）と、判定された所定のオブジェクトとに基づいて、判定された所定のオブジェクトに対応する命令（または、命令群）を特定する。そして、本実施形態に係る情報処理装置は、取得された音声信号から、特定された命令を認識するこ

50

とによって、所定のオブジェクトに対応する命令を認識する。

【0102】

(A-2)

また、認識する命令を動的に変える制御の制御対象が上記外部装置である場合、本実施形態に係る情報処理装置は、例えば、“認識する命令を動的に変えさせる命令”と、所定のオブジェクトに対応するオブジェクトを示す情報とを含む制御データを、通信部（後述する）などに、上記外部装置へと送信させる。本実施形態に係るオブジェクトを示す情報としては、例えば、オブジェクトを示すIDや、オブジェクトを示すデータなどが挙げられる。また、上記制御データには、例えば、さらにユーザが発話した音声を示す音声信号が含まれていてもよい。上記制御データを取得した上記外部装置は、例えば、上記（A-1）に示す本実施形態に係る情報処理装置と同様の処理を行うことによって、所定のオブジェクトに対応する命令を認識する。

10

【0103】

(B) 本実施形態に係る音声認識に係る処理における、認識する命令の動的な変更の第2の例

本実施形態に係る情報処理装置は、上記（1）の処理（判定処理）においてユーザが見たと判定された所定のオブジェクトを含む表示画面における領域内に含まれる、他のオブジェクトに対応する命令を認識するように制御する。また、本実施形態に係る情報処理装置は、例えば、上記（A）に示すように所定のオブジェクトに対応する命令を認識することに加え、さらに（B）の処理を行ってもよい。

20

【0104】

ここで、本実施形態に係る所定のオブジェクトを含む表示画面における領域としては、例えば、本実施形態に係る第1領域よりも大きな領域が挙げられる。一例を挙げると、本実施形態に係る所定のオブジェクトを含む表示画面における領域としては、例えば、所定のオブジェクトの基準点を中心とする円形や矩形の領域や、分割領域などが挙げられる。

【0105】

(B-1)

認識する命令を動的に変える制御の制御対象が自装置である場合、本実施形態に係る情報処理装置は、例えば、本実施形態に係る所定のオブジェクトを含む表示画面における領域に基準位置が含まれるオブジェクトのうち、所定のオブジェクト以外のオブジェクトを、他のオブジェクトと判定する。なお、本実施形態に係る他のオブジェクトの判定方法は、上記に限られない。例えば、本実施形態に係る情報処理装置は、本実施形態に係る所定のオブジェクトを含む表示画面における領域内に少なくとも一部が表示されるオブジェクトのうち、所定のオブジェクト以外のオブジェクトを、他のオブジェクトとして判定してもよい。

30

【0106】

また、本実施形態に係る情報処理装置は、例えば、オブジェクトと命令（または、命令群）とが対応付けられているテーブル（または、データベース）と、判定された他のオブジェクトとに基づいて、他のオブジェクトに対応する命令（または、命令群）を特定する。本実施形態に係る情報処理装置は、例えば、上記テーブル（または、データベース）と、判定された所定のオブジェクトとに基づいて、判定された所定のオブジェクトに対応する命令（または、命令群）をさらに特定してもよい。そして、本実施形態に係る情報処理装置は、取得された音声信号から、特定された命令を認識することによって、他のオブジェクトに対応する命令（または、さらに所定のオブジェクトに対応する命令）を認識する。

40

【0107】

(B-2)

また、認識する命令を動的に変える制御の制御対象が上記外部装置である場合、本実施形態に係る情報処理装置は、例えば、“認識する命令を動的に変えさせる命令”と、他のオブジェクトに対応するオブジェクトを示す情報とを含む制御データを、通信部（後述す

50

る)などに、上記外部装置へと送信させる。また、上記制御データには、例えば、さらにユーザが発話した音声を示す音声信号や、所定のオブジェクトに対応するオブジェクトを示す情報が含まれていてもよい。上記制御データを取得した上記外部装置は、例えば、上記(B-1)に示す本実施形態に係る情報処理装置と同様の処理を行うことによって、他のオブジェクトに対応する命令に対応する命令(または、さらに所定のオブジェクトに対応する命令)を認識する。

【0108】

本実施形態に係る情報処理装置は、本実施形態に係る音声認識制御処理として、例えば上記のような処理を行う。

【0109】

なお、本実施形態に係る音声認識制御処理は、上記に示す処理に限られない。

【0110】

例えば、上記(1)の処理(判定処理)において、ユーザが所定のオブジェクトを見たと判定した後に、ユーザが所定のオブジェクトを見ていないと判定された場合には、本実施形態に係る情報処理装置は、所定のオブジェクトを見たと判定されたユーザに対する音声認識を終了させる。

【0111】

本実施形態に係る情報処理装置は、本実施形態に係る情報処理方法に係る処理として、例えば、上記(1)の処理(判定処理)、および上記(2)の処理(音声認識制御処理)を行う。

【0112】

ここで、本実施形態に係る情報処理装置は、上記(1)の処理(判定処理)において所定のオブジェクトを見たと判定したときに、上記(2)の処理(音声認識制御処理)を行う。つまり、ユーザは、例えば、所定のオブジェクトに視線を向けて所定のオブジェクトを見ることによって、本実施形態に係る情報処理装置に音声認識を開始させることが可能となる。また、上述したように、仮に、ユーザが、他の操作や会話などを行っていたとしても、ユーザが所定のオブジェクトを見ることにより当該他の操作や会話を妨げる可能性は、特定のユーザ操作や特定のワードの発話によって音声認識が行われる場合よりも低い。また、上述したように、ユーザが所定のオブジェクトを見ることは、上記特定のユーザ操作や特定のワードの発話よりも、より自然な操作であるといえる。

【0113】

したがって、本実施形態に係る情報処理装置が、本実施形態に係る情報処理方法に係る処理として、上記(1)の処理(判定処理)、および上記(2)の処理(音声認識制御処理)を行うことによって、音声認識が行われる場合におけるユーザの利便性の向上を図ることができる。

【0114】

なお、本実施形態に係る情報処理方法に係る処理は、上記(1)の処理(判定処理)、および上記(2)の処理(音声認識制御処理)に限られない。

【0115】

例えば、本実施形態に係る情報処理装置は、本実施形態に係る所定のオブジェクトを表示画面に表示させる処理(表示制御処理)を行うことも可能である。そこで、次に、本実施形態に係る表示制御処理について説明する。

【0116】

(3)表示制御処理

本実施形態に係る情報処理装置は、本実施形態に係る所定のオブジェクトを表示画面に表示させる。より具体的には、本実施形態に係る情報処理装置は、例えば、下記の第1の例に係る表示制御処理～下記の第4の例に係る表示制御処理の処理を行う。

【0117】

(3-1)表示制御処理の第1の例

本実施形態に係る情報処理装置は、例えば、表示画面における設定されている位置に、

10

20

30

40

50

所定のオブジェクトを表示させる。つまり、本実施形態に係る情報処理装置は、ユーザの視線の位置に関する情報が示す視線の位置がどのような位置であっても、ユーザの視線の位置に関する情報が示す視線の位置によらずに、設定されている位置に、所定のオブジェクトを表示させる。

【0118】

本実施形態に係る情報処理装置は、例えば、所定のオブジェクトを常に表示画面に表示させる。なお、本実施形態に係る情報処理装置は、例えば、視線による操作以外のユーザ操作に基づいて、所定のオブジェクトを選択的に表示させることも可能である。

【0119】

図5は、本実施形態に係る情報処理方法に係る処理の一例を説明するための説明図であり、本実施形態に係る表示制御処理により表示される所定のオブジェクトOの表示位置の一例を示している。図5では、所定のオブジェクトOが音声認識アイコンである例を示している。

10

【0120】

所定のオブジェクトが表示される位置の一例としては、例えば、図5のAに示すような表示画面の画面端の位置や、図5のBに示すような表示画面の中央の位置、図1において符号O1～O3で表されるオブジェクトが表示されている位置など様々な位置が挙げられる。なお、所定のオブジェクトが表示される位置は、図1、図5に示す例に限られず、表示画面の任意の位置であってもよい。

【0121】

20

(3-2) 表示制御処理の第2の例

本実施形態に係る情報処理装置は、例えば、ユーザの視線の位置に関する情報に基づいて、所定のオブジェクトを選択的に表示させる。

【0122】

より具体的には、本実施形態に係る情報処理装置は、例えば、ユーザの視線の位置に関する情報が示す視線の位置が、設定されている領域に含まれる場合に、所定のオブジェクトを表示させる。ユーザの視線の位置に関する情報が示す視線の位置が設定されている領域に含まれるときに、所定のオブジェクトが表示される場合には、ユーザが、設定されている領域を一度見ることによって、所定のオブジェクトが表示される。

【0123】

30

ここで、本実施形態に係る表示制御処理における上記領域としては、例えば、所定のオブジェクトを含む領域のうちの最小の領域(すなわち、所定のオブジェクトが表示される領域)や、所定のオブジェクトの基準点を中心とする円形や矩形の領域、分割領域などが挙げられる。

【0124】

なお、第2の例に係る表示制御処理は、上記に限られない。

【0125】

例えば、所定のオブジェクトを表示させる場合、本実施形態に係る情報処理装置は、ユーザの視線の位置に関する情報が示す視線の位置に基づいて、段階的に所定のオブジェクトを表示させてもよい。本実施形態に係る情報処理装置は、例えば、ユーザの視線の位置に関する情報が示す視線の位置が設定されている領域に含まれる時間に応じて、段階的に所定のオブジェクトを表示させる。

40

【0126】

図6は、本実施形態に係る情報処理方法に係る処理の一例を説明するための説明図であり、本実施形態に係る表示制御処理により段階的に表示される所定のオブジェクトOの一例を示している。図6では、所定のオブジェクトOが音声認識アイコンである例を示している。

【0127】

例えば、本実施形態に係る情報処理装置は、ユーザの視線の位置に関する情報が示す視線の位置が設定されている領域に含まれる時間が、第1の時間以上である場合(または、

50

当該設定されている領域に含まれる時間が、第 1 の時間より大きい場合) に、所定のオブジェクト O の一部を表示画面に表示させる (図 6 に示す A) 。本実施形態に係る情報処理装置は、例えば、ユーザの視線の位置に関する情報が示す視線の位置に対応する位置に、所定のオブジェクト O の一部を表示させる。

【 0 1 2 8 】

ここで、本実施形態に係る第 1 の時間としては、例えば、設定された固定の時間が挙げられる。

【 0 1 2 9 】

また、本実施形態に係る情報処理装置は、取得されているユーザの視線の位置に関する情報の数 (すなわち、ユーザ数) に基づいて、第 1 の時間を動的に変えてもよい。本実施形態に係る情報処理装置は、例えば、ユーザ数が多い程、第 1 の時間を長く設定する。第 1 の時間がユーザ数に応じて動的に設定されることによって、例えば、一のユーザが、偶然、所定のオブジェクトを表示させてしまうことを防止することができる。

【 0 1 3 0 】

例えば図 6 の A に示すように所定のオブジェクト O の一部が表示画面に表示されると、本実施形態に係る情報処理装置は、所定のオブジェクト O の一部が表示画面に表示されてからのユーザの視線の位置に関する情報が示す視線の位置が設定されている領域に含まれる時間が、第 2 の時間以上である場合 (または、当該設定されている領域に含まれる時間が、第 2 の時間より大きい場合) に、所定のオブジェクト O の全体を表示画面に表示させる (図 6 に示す B) 。

【 0 1 3 1 】

ここで、本実施形態に係る第 2 の時間としては、例えば、設定された固定の時間が挙げられる。

【 0 1 3 2 】

また、本実施形態に係る情報処理装置は、上記第 1 の時間と同様に、取得されているユーザの視線の位置に関する情報の数 (すなわち、ユーザ数) に基づいて、第 2 の時間を動的に変えてもよい。第 2 の時間がユーザ数に応じて動的に設定されることによって、例えば、一のユーザが、偶然、所定のオブジェクトを表示させてしまうことを防止することができる。

【 0 1 3 3 】

また、所定のオブジェクトを表示させる場合、本実施形態に係る情報処理装置は、例えば、設定されている表示方法を用いて、所定のオブジェクトを表示させてもよい。

【 0 1 3 4 】

本実施形態に係る設定されている表示方法としては、例えば、スライドインや、フェードインなどが挙げられる。

【 0 1 3 5 】

また、本実施形態に係る情報処理装置は、例えば、ユーザの視線の位置に関する情報に基づいて、本実施形態に係る設定されている表示方法を動的に変えることも可能である。

【 0 1 3 6 】

一例を挙げると、本実施形態に係る情報処理装置は、ユーザの視線の位置に関する情報に基づいて目の動きの方向 (例えば、上下や左右など) を特定する。そして、本実施形態に係る情報処理装置は、特定された目の動きの方向に対応する方向から、所定のオブジェクトが現れるような表示方法を用いて、所定のオブジェクトを表示させる。また、本実施形態に係る情報処理装置は、例えば、さらに、ユーザの視線の位置に関する情報が示す視線の位置に応じて、所定のオブジェクトが現れる位置を変えてもよい。

【 0 1 3 7 】

(3 - 3) 表示制御処理の第 3 の例

本実施形態に係る情報処理装置は、例えば、上記 (2) の処理 (音声認識制御処理) により音声認識が行われている場合、所定のオブジェクトの表示態様を変える。本実施形態に係る情報処理装置が所定のオブジェクトの表示態様を変えることによって、本実施形態

10

20

30

40

50

に係る情報処理方法に係る処理の状態を、ユーザに対してフィードバックすることが可能となる。

【0138】

図7は、本実施形態に係る情報処理方法に係る処理の一例を説明するための説明図であり、本実施形態に係る所定のオブジェクトの表示態様の一例を示している。図7のA～図7Eは、本実施形態に係る所定のオブジェクトの表示態様の一例をそれぞれ示している。

【0139】

本実施形態に係る情報処理装置は、例えば、図7のAに示すように、上記(1)の処理(判定処理)において所定のオブジェクトを見たときと判定されたユーザに応じて、所定のオブジェクトの色や所定のオブジェクトが光る色を変える。所定のオブジェクトの色や所定のオブジェクトが光る色が変わることによって、上記(1)の処理(判定処理)において所定のオブジェクトを見たときと判定されたユーザを、表示画面をみている1または2以上のユーザに対してフィードバックすることができる。

10

【0140】

ここで、本実施形態に係る情報処理装置は、例えば、上記(1)の処理(判定処理)においてユーザIDを認識する場合には、ユーザIDに対応する色の所定のオブジェクトや、ユーザIDに対応する色で光る所定のオブジェクトを表示させる。また、本実施形態に係る情報処理装置は、例えば、上記(1)の処理(判定処理)において所定のオブジェクトを見たときと判定されるごとに、異なる色の所定のオブジェクトや、異なる色で光る所定のオブジェクトを表示させてもよい。

20

【0141】

また、本実施形態に係る情報処理装置は、例えば、図7のBや図7のCに示すように、上記(2)の処理(音声認識制御処理)において認識した音声の方向を視覚的に示してもよい。認識した音声の方向を視覚的に示すことによって、本実施形態に係る情報処理装置が認識した音声の方向を、表示画面をみている1または2以上のユーザに対してフィードバックすることができる。

【0142】

図7のBに示す例では、図7のBに示す符号DIで示されているように、音声の方向部分が空いたバーによって、認識した音声の方向を示している。また、図7のCに示す例では、キャラクター画像(音声認識画像の一例)が認識した音声の方向を見ることによって、認識した音声の方向を示している。

30

【0143】

また、本実施形態に係る情報処理装置は、例えば、図7のDや図7のEに示すように、上記(1)の処理(判定処理)において所定のオブジェクトを見たときと判定されたユーザに対応する撮像画像を、音声認識アイコンと併せて示してもよい。撮像画像を音声認識アイコンと併せて示すことによって、上記(1)の処理(判定処理)において所定のオブジェクトを見たときと判定されたユーザを、表示画面をみている1または2以上のユーザに対してフィードバックすることができる。

【0144】

図7のDに示す例は、撮像画像が音声認識アイコンと並んで表示された例を示している。また、図7のEに示す例は、撮像画像を音声認識アイコンに合成して表示された例を示している。

40

【0145】

本実施形態に係る情報処理装置は、例えば図7に示すように、所定のオブジェクトの表示態様を変えることによって、本実施形態に係る情報処理方法に係る処理の状態を、ユーザに対してフィードバックする。

【0146】

なお、第3の例に係る表示制御処理は、図7に示す例に限られない。例えば、本実施形態に係る情報処理装置は、上記(1)の処理(判定処理)においてユーザIDを認識する場合には、ユーザIDに対応するオブジェクト(例えば、音声認識アイコンや、キャラク

50

ター画像などの音声認識画像など)を表示させてもよい。

【0147】

(3-4)表示制御処理の第4の例

本実施形態に係る情報処理装置は、例えば、上記第1の例に係る表示制御処理または第2の例に係る表示制御処理と、第3の例に係る表示制御処理とを組み合わせ処理を行うことも可能である。

【0148】

(本実施形態に係る情報処理装置)

次に、上述した本実施形態に係る情報処理方法に係る処理を行うことが可能な、本実施形態に係る情報処理装置の構成の一例について、説明する。

【0149】

図8は、本実施形態に係る情報処理装置100の構成の一例を示すブロック図である。情報処理装置100は、例えば、通信部102と、制御部104とを備える。

【0150】

また、情報処理装置100は、例えば、ROM(Read Only Memory。図示せず)や、RAM(Random Access Memory。図示せず)、記憶部(図示せず)、ユーザが操作可能な操作部(図示せず)、様々な画面を表示画面に表示する表示部(図示せず)などを備えていてもよい。情報処理装置100は、例えば、データの伝送路としてのバス(bus)により上記各構成要素間を接続する。

【0151】

ここで、ROM(図示せず)は、制御部104が使用するプログラムや演算パラメータなどの制御用データを記憶する。RAM(図示せず)は、制御部104により実行されるプログラムなどを一時的に記憶する。

【0152】

記憶部(図示せず)は、情報処理装置100が備える記憶手段であり、例えば、表示画面に表示される各種オブジェクトを示すデータなどの本実施形態に係る情報処理方法に係るデータや、アプリケーションなど様々なデータを記憶する。ここで、記憶部(図示せず)としては、例えば、ハードディスク(Hard Disk)などの磁気記録媒体や、フラッシュメモリ(flash memory)などの不揮発性メモリ(nonvolatile memory)などが挙げられる。また、記憶部(図示せず)は、情報処理装置100から着脱可能であってもよい。

【0153】

操作部(図示せず)としては、後述する操作入力デバイスが挙げられる。また、表示部(図示せず)としては、後述する表示デバイスが挙げられる。

【0154】

[情報処理装置100のハードウェア構成例]

図9は、本実施形態に係る情報処理装置100のハードウェア構成の一例を示す説明図である。情報処理装置100は、例えば、MPU150と、ROM152と、RAM154と、記録媒体156と、入出力インタフェース158と、操作入力デバイス160と、表示デバイス162と、通信インタフェース164とを備える。また、情報処理装置100は、例えば、データの伝送路としてのバス166で各構成要素間を接続する。

【0155】

MPU150は、例えば、MPU(Micro Processing Unit)などのプロセッサや各種処理回路などで構成され、情報処理装置100全体を制御する制御部104として機能する。また、MPU150は、情報処理装置100において、例えば、後述する判定部110、音声認識制御部112、および表示制御部114の役目を果たす。

【0156】

ROM152は、MPU150が使用するプログラムや演算パラメータなどの制御用データなどを記憶する。RAM154は、例えば、MPU150により実行されるプログラムなどを一時的に記憶する。

【0157】

10

20

30

40

50

記録媒体 156 は、記憶部（図示せず）として機能し、例えば、表示画面に表示される各種オブジェクトを示すデータなどの本実施形態に係る情報処理方法に係るデータや、アプリケーションなど様々なデータを記憶する。ここで、記録媒体 156 としては、例えば、ハードディスクなどの磁気記録媒体や、フラッシュメモリなどの不揮発性メモリが挙げられる。また、記録媒体 156 は、情報処理装置 100 から着脱可能であってもよい。

【0158】

入出力インタフェース 158 は、例えば、操作入力デバイス 160 や、表示デバイス 162 を接続する。操作入力デバイス 160 は、操作部（図示せず）として機能し、また、表示デバイス 162 は、表示部（図示せず）として機能する。ここで、入出力インタフェース 158 としては、例えば、USB（Universal Serial Bus）端子や、DVI（Digital Visual Interface）端子、HDMI（High-Definition Multimedia Interface）（登録商標）端子、各種処理回路などが挙げられる。また、操作入力デバイス 160 は、例えば、情報処理装置 100 上に備えられ、情報処理装置 100 の内部で入出力インタフェース 158 と接続される。操作入力デバイス 160 としては、例えば、ボタンや、方向キー、ジョグダイヤルなどの回転型セレクター、あるいは、これらの組み合わせなどが挙げられる。また、表示デバイス 162 は、例えば、情報処理装置 100 上に備えられ、情報処理装置 100 の内部で入出力インタフェース 158 と接続される。表示デバイス 162 としては、例えば、液晶ディスプレイ（Liquid Crystal Display）や有機 EL ディスプレイ（Organic Electro-Luminescence Display。または、OLED ディスプレイ（Organic Light Emitting Diode Display）ともよばれる。）などが挙げられる。

【0159】

なお、入出力インタフェース 158 が、情報処理装置 100 の外部装置としての操作入力デバイス（例えば、キーボードやマウスなど）や表示デバイスなどの、外部デバイスと接続することも可能であることは、言うまでもない。また、表示デバイス 162 は、例えばタッチスクリーンなど、表示とユーザ操作とが可能なデバイスであってもよい。

【0160】

通信インタフェース 164 は、情報処理装置 100 が備える通信手段であり、ネットワークを介して（あるいは、直接的に）、外部の撮像デバイスや、外部の表示デバイス、外部のセンサなどの、外部デバイスや外部装置と、無線または有線で通信を行うための通信部 102 として機能する。ここで、通信インタフェース 164 としては、例えば、通信アンテナおよび RF（Radio Frequency）回路（無線通信）や、IEEE 802.15.1 ポートおよび送受信回路（無線通信）、IEEE 802.11 ポートおよび送受信回路（無線通信）、あるいは LAN（Local Area Network）端子および送受信回路（有線通信）などが挙げられる。また、本実施形態に係るネットワークとしては、例えば、LAN や WAN（Wide Area Network）などの有線ネットワーク、無線 LAN（WLAN: Wireless Local Area Network）や基地局を介した無線 WAN（WWAN: Wireless Wide Area Network）などの無線ネットワーク、あるいは、TCP/IP（Transmission Control Protocol/Internet Protocol）などの通信プロトコルを用いたインターネットなどが挙げられる。

【0161】

情報処理装置 100 は、例えば図 9 に示す構成によって、本実施形態に係る情報処理方法に係る処理を行う。なお、本実施形態に係る情報処理装置 100 のハードウェア構成は、図 9 に示す構成に限られない。

【0162】

例えば、情報処理装置 100 は、動画画像または静止画像を撮像する撮像部（図示せず）の役目を果たす撮像デバイスを備えていてもよい。撮像デバイスを備える場合には、情報処理装置 100 は、例えば、撮像デバイスにおいて撮像により生成された撮像画像を処理して、ユーザの視線の位置に関する情報を得ることが可能となる。また、撮像デバイスを備える場合には、情報処理装置 100 は、例えば、撮像デバイスにおいて撮像により生成された撮像画像を用いたユーザの特定に係る処理を行うことや、当該撮像画像（または、

撮像画像の一部)を、オブジェクトとして用いることも可能である。

【0163】

ここで、本実施形態に係る撮像デバイスとしては、例えば、レンズ/撮像素子と信号処理回路とが挙げられる。レンズ/撮像素子は、例えば、光学系のレンズと、CMOS (Complementary Metal Oxide Semiconductor) などの撮像素子を複数用いたイメージセンサとで構成される。信号処理回路は、例えば、AGC (Automatic Gain Control) 回路やADC (Analog to Digital Converter) を備え、撮像素子により生成されたアナログ信号をデジタル信号 (画像データ) に変換する。また、信号処理回路は、例えば、White Balance 補正処理や、色調補正処理、ガンマ補正処理、YCbCr 変換処理、エッジ強調処理など各種信号処理を行ってもよい。

10

【0164】

また、情報処理装置100は、例えば、本実施形態に係るユーザの視線の位置の特定に用いることが可能なデータを得る検出部 (図示せず) の役目を果たすセンサをさらに備えていてもよい。センサを備える場合には、情報処理装置100は、例えば、センサから得られるデータを用いて、ユーザの視線の位置の推定精度の向上を図ることが可能となる。

【0165】

本実施形態に係るセンサとしては、例えば、赤外線センサなど、ユーザの視線の位置の推定精度の向上に用いることが可能な検出値を得る任意のセンサが挙げられる。

【0166】

また、情報処理装置100は、例えば、スタンドアロンで処理を行う構成である場合には、通信インタフェース164を備えていなくてもよい。また、情報処理装置100は、記録媒体156や、操作デバイス160、表示デバイス162を備えない構成をとることも可能である。

20

【0167】

再度図8を参照して、情報処理装置100の構成の一例について説明する。通信部102は、情報処理装置100が備える通信手段であり、ネットワークを介して (あるいは、直接的に)、外部の撮像デバイスや、外部の表示デバイス、外部のセンサなどの、外部デバイスや外部装置と、無線または有線で通信を行う。また、通信部102は、例えば制御部104により通信が制御される。

【0168】

ここで、通信部102としては、例えば、通信アンテナおよびRF回路や、LAN端子および送受信回路などが挙げられるが、通信部102の構成は、上記に限られない。例えば、通信部102は、USB端子および送受信回路など通信を行うことが可能な任意の規格に対応する構成や、ネットワークを介して外部装置と通信可能な任意の構成をとることができる。

30

【0169】

制御部104は、例えばMPUなどで構成され、情報処理装置100全体を制御する役目を果たす。また、制御部104は、例えば、判定部110と、音声認識制御部112と、表示制御部114とを備え、本実施形態に係る情報処理方法に係る処理を主導的に行う役目を果たす。

40

【0170】

判定部110は、上記(1)の処理 (判定処理) を主導的に行う役目を果たす。

【0171】

例えば、判定部110は、ユーザの視線の位置に関する情報に基づいて、ユーザが所定のオブジェクトを見たかを判定する。より具体的には、判定部110は、例えば、上記(1-1)に示す第1の例に係る判定処理を行う。

【0172】

また、判定部110は、例えば、ユーザの視線の位置に関する情報に基づいて、ユーザが所定のオブジェクトを見たか判定された後に、ユーザが所定のオブジェクトを見ていないことを判定することも可能である。より具体的には、判定部110は、例えば、上記(

50

１－２）に示す第２の例に係る判定処理や、上記（１－３）に示す第３の例に係る判定処理を行う。

【０１７３】

また、判定部１１０は、例えば、上記（１－４）に示す第４の例に係る判定処理や、上記（１－５）に示す第５の例に係る判定処理を行ってもよい。

【０１７４】

音声認識制御部１１２は、上記（２）の処理（音声認識制御処理）を主導的に行う役目を果たす。

【０１７５】

例えば、音声認識制御部１１２は、判定部１１０においてユーザが所定のオブジェクトを見たときと判定された場合に、音声認識処理を制御し、音声認識を行わせる。より具体的には、音声認識制御部１１２は、例えば、上記（２－１）に示す第１の例に係る音声認識制御処理や、上記（２－２）に示す第２の例に係る音声認識制御処理を行う。

10

【０１７６】

また、音声認識制御部１１２は、判定部１１０において、ユーザが所定のオブジェクトを見たときと判定した後、ユーザが所定のオブジェクトを見ていないと判定された場合には、所定のオブジェクトを見たときと判定されたユーザに対する音声認識を終了させる。

【０１７７】

表示制御部１１４は、上記（３）の処理（表示制御処理）を主導的に行う役目を果たし、本実施形態に係る所定のオブジェクトを表示画面に表示させる。より具体的には、表示制御部１１４は、例えば、上記（３－１）に示す第１の例に係る表示制御処理や、上記（３－２）に示す第２の例に係る表示制御処理、上記（３－３）に示す第３の例に係る表示制御処理を行う。

20

【０１７８】

制御部１０４は、例えば、判定部１１０、音声認識制御部１１２、および表示制御部１１４を備えることによって、本実施形態に係る情報処理方法に係る処理を主導的に行う。

【０１７９】

情報処理装置１００は、例えば図８に示す構成によって、本実施形態に係る情報処理方法に係る処理（例えば、上記（１）の処理（判定処理）～上記（３）の処理（表示制御処理））を行う。

30

【０１８０】

したがって、情報処理装置１００は、例えば図８に示す構成によって、音声認識が行われる場合におけるユーザの利便性の向上を図ることができる。

【０１８１】

また、例えば図８に示す構成によって、情報処理装置１００は、例えば上述したような、本実施形態に係る情報処理方法に係る処理が行われることにより奏される効果を奏することができる。

【０１８２】

なお、本実施形態に係る情報処理装置の構成は、図８に示す構成に限られない。

【０１８３】

例えば、本実施形態に係る情報処理装置は、図８に示す判定部１１０、音声認識制御部１１２、および表示制御部１１４のうちの１または２以上を、制御部１０４とは個別に備える（例えば、個別の処理回路で実現する）ことができる。

40

【０１８４】

また、本実施形態に係る情報処理装置は、例えば、図８に示す表示制御部１１４を備えない構成をとることも可能である。表示制御部１１４を備えない構成であっても、本実施形態に係る情報処理装置は、上記（１）の処理（判定処理）、および上記（２）の処理（音声認識制御処理）を行うことが可能である。よって、表示制御部１１４を備えない構成であっても、本実施形態に係る情報処理装置は、音声認識が行われる場合におけるユーザの利便性の向上を図ることができる。

50

【0185】

また、例えば、通信部102と同様の機能、構成を有する外部の通信デバイスを介して、外部のデバイスや外部装置と通信を行う場合や、スタンドアロンで処理を行う構成である場合には、本実施形態に係る情報処理装置は、通信部102を備えていなくてもよい。

【0186】

また、本実施形態に係る情報処理装置は、例えば、撮像デバイスで構成される撮像部（図示せず）を、さらに備えていてもよい。撮像部（図示せず）を備える場合には、本実施形態に係る情報処理装置は、例えば、撮像部（図示せず）において撮像により生成された撮像画像を処理して、ユーザの視線の位置に関する情報を得ることが可能となる。また、撮像部（図示せず）を備える場合には、本実施形態に係る情報処理装置は、例えば、撮像部（図示せず）において撮像により生成された撮像画像を用いたユーザの特定に係る処理を行うことや、当該撮像画像（または、撮像画像の一部）を、オブジェクトとして用いることも可能である。

10

【0187】

また、本実施形態に係る情報処理装置は、例えば、ユーザの視線の位置の推定精度の向上に用いることが可能な検出値を得る任意のセンサで構成される検出部（図示せず）を、さらに備えていてもよい。検出部（図示せず）を備える場合には、本実施形態に係る情報処理装置は、例えば、検出部（図示せず）から得られるデータを用いて、ユーザの視線の位置の推定精度の向上を図ることが可能となる。

【0188】

20

以上、本実施形態として、情報処理装置を挙げて説明したが、本実施形態は、かかる形態に限られない。本実施形態は、例えば、テレビ受像機や、表示装置、タブレット型の装置、携帯電話やスマートフォンなどの通信装置、映像/音楽再生装置（または映像/音楽記録再生装置）、ゲーム機、PC（Personal Computer）などのコンピュータなど、様々な機器に適用することができる。また、本実施形態は、例えば、上記のような機器に組み込むことが可能な、処理IC（Integrated Circuit）に適用することもできる。

【0189】

また、本実施形態は、例えばクラウドコンピューティングなどのように、ネットワークへの接続（または各装置間の通信）を前提とした、複数の装置からなるシステムにより実現されてもよい。つまり、上述した本実施形態に係る情報処理装置は、例えば、複数の装置からなる情報処理システムとして実現することも可能である。

30

【0190】

（本実施形態に係るプログラム）

コンピュータを、本実施形態に係る情報処理装置として機能させるためのプログラム（例えば、“上記（1）の処理（判定処理）、および上記（2）の処理（音声認識制御処理）”や“上記（1）の処理（判定処理）～（3）の処理（表示制御処理）”など、本実施形態に係る情報処理方法に係る処理を実行することが可能なプログラム）が、コンピュータにおいてプロセッサなどにより実行されることによって、音声認識が行われる場合におけるユーザの利便性の向上を図ることができる。

【0191】

40

また、コンピュータを、本実施形態に係る情報処理装置として機能させるためのプログラムが、コンピュータにおいてプロセッサなどにより実行されることによって、上述した本実施形態に係る情報処理方法に係る処理によって奏される効果を奏することができる。

【0192】

以上、添付図面を参照しながら本開示の好適な実施形態について詳細に説明したが、本開示の技術的範囲はかかる例に限定されない。本開示の技術分野における通常の知識を有する者であれば、特許請求の範囲に記載された技術的思想の範疇内において、各種の変更例または修正例に想到し得ることは明らかであり、これらについても、当然に本開示の技術的範囲に属するものと了解される。

【0193】

50

例えば、上記では、コンピュータを、本実施形態に係る情報処理装置として機能させるためのプログラム（コンピュータプログラム）が提供されることを示したが、本実施形態は、さらに、上記プログラムを記憶させた記録媒体も併せて提供することができる。

【0194】

上述した構成は、本実施形態の一例を示すものであり、当然に、本開示の技術的範囲に属するものである。

【0195】

また、本明細書に記載された効果は、あくまで説明的または例示的なものであって限定的ではない。つまり、本開示に係る技術は、上記の効果とともに、または上記の効果に代えて、本明細書の記載から当業者には明らかな他の効果を奏しうる。

【0196】

なお、以下のような構成も本開示の技術的範囲に属する。

(1)

表示画面におけるユーザの視線の位置に関する情報に基づいて、ユーザが所定のオブジェクトを見たかを判定する判定部と、

ユーザが所定のオブジェクトを見たと判定された場合に、音声認識処理を制御する音声認識制御部と、

を備える、情報処理装置。

(2)

前記音声認識制御部は、見たと判定された前記所定のオブジェクトに基づいて、認識する命令を動的に変えるよう制御する、(1)に記載の情報処理装置。

(3)

前記音声認識制御部は、見たと判定された前記所定のオブジェクトに対応する命令を認識するよう制御する、(1)、または(2)に記載の情報処理装置。

(4)

前記音声認識制御部は、見たと判定された前記所定のオブジェクトを含む表示画面における領域内に含まれる、他のオブジェクトに対応する命令を認識するよう制御する、(1)~(3)のいずれか1つに記載の情報処理装置。

(5)

前記音声認識制御部は、

所定のオブジェクトを見たと判定されたユーザに対応するユーザの視線の位置に関する情報に基づいて、音源分離を行うことが可能な音声入力デバイスに、所定のオブジェクトを見たと判定されたユーザの位置から発せられる音声を示す音声信号を取得させ、

前記音声入力デバイスにより取得された音声信号に対して音声認識を行わせる、(1)~(4)のいずれか1つに記載の情報処理装置。

(6)

前記音声認識制御部は、

所定のオブジェクトを見たと判定されたユーザに対応するユーザの視線の位置に関する情報に基づくユーザの位置と、音源定位を行うことが可能な音声入力デバイスが測定した音源の位置との差分が、設定された閾値以下の場合、または、

前記ユーザの位置と前記音源の位置との差分が、前記閾値より小さい場合に、

前記音声入力デバイスにより取得された音声を示す音声信号に対して音声認識を行わせる、(1)~(4)のいずれか1つに記載の情報処理装置。

(7)

前記判定部は、ユーザの視線の位置に関する情報が示す視線の位置が、所定のオブジェクトを含む表示画面における第1領域内に含まれる場合に、ユーザが所定のオブジェクトを見たと判定する、(1)~(6)のいずれか1つに記載の情報処理装置。

(8)

前記判定部が、ユーザが所定のオブジェクトを見たと判定した場合、

前記判定部は、所定のオブジェクトを見たと判定されたユーザに対応するユーザの視線

10

20

30

40

50

の位置に関する情報が示す視線の位置が、所定のオブジェクトを含む表示画面における第2領域内に含まれなくなったときに、前記ユーザが所定のオブジェクトを見ていないと判定し、

前記音声認識制御部は、前記ユーザが所定のオブジェクトを見ていないと判定されたときに、前記ユーザに対する音声認識を終了させる、(1)~(7)のいずれか1つに記載の情報処理装置。

(9)

前記判定部が、ユーザが所定のオブジェクトを見たと判定した場合、
前記判定部は、

所定のオブジェクトを見たと判定されたユーザに対応するユーザの視線の位置に関する情報が示す視線の位置が、所定のオブジェクトを含む表示画面における第2領域内に含まれない状態が、設定された設定時間以上継続するとき、または、

所定のオブジェクトを見たと判定されたユーザに対応するユーザの視線の位置に関する情報が示す視線の位置が前記第2領域内に含まれない状態が、前記設定時間より長く継続するとき、

前記ユーザが所定のオブジェクトを見ていないと判定し、

前記音声認識制御部は、前記ユーザが所定のオブジェクトを見ていないと判定されたときに、前記ユーザに対する音声認識を終了させる、(1)~(7)のいずれか1つに記載の情報処理装置。

(10)

前記判定部は、所定のオブジェクトを見たと判定されたユーザに対応するユーザの視線の位置に関する情報が示す視線の位置の履歴に基づいて、前記設定時間を動的に設定する、(9)に記載の情報処理装置。

(11)

前記判定部は、一のユーザが所定のオブジェクトを見たと判定した後に、前記一のユーザが所定のオブジェクトを見ていないと判定されていない場合には、他のユーザが所定のオブジェクトを見たとは判定しない、(1)~(10)のいずれか1つに記載の情報処理装置。

(12)

前記判定部は、

前記表示画面において画像が表示される方向が撮像された撮像画像に基づいてユーザを特定し、

特定されたユーザに対応するユーザの視線の位置に関する情報に基づいて、ユーザが所定のオブジェクトを見たかを判定する、(1)~(11)のいずれか1つに記載の情報処理装置。

(13)

前記所定のオブジェクトを表示画面に表示させる表示制御部をさらに備える、(1)~(12)のいずれか1つに記載の情報処理装置。

(14)

前記表示制御部は、前記ユーザの視線の位置に関する情報が示す視線の位置によらず、表示画面における設定されている位置に、前記所定のオブジェクトを表示させる、(13)に記載の情報処理装置。

(15)

前記表示制御部は、前記ユーザの視線の位置に関する情報に基づいて、前記所定のオブジェクトを選択的に表示させる、(13)に記載の情報処理装置。

(16)

前記表示制御部は、前記所定のオブジェクトを表示させる場合には、設定されている表示方法を用いて前記所定のオブジェクトを表示させる、(15)に記載の情報処理装置。

(17)

前記表示制御部は、前記所定のオブジェクトを表示させる場合には、前記ユーザの視線

10

20

30

40

50

の位置に関する情報が示す視線の位置に基づいて、段階的に前記所定のオブジェクトを表示させる、(15)、または(16)に記載の情報処理装置。

(18)

前記表示制御部は、音声認識が行われている場合、前記所定のオブジェクトの表示態様を変える、(13)～(17)のいずれか1つに記載の情報処理装置。

(19)

表示画面におけるユーザの視線の位置に関する情報に基づいて、ユーザが所定のオブジェクトを見たかを判定するステップと、

ユーザが所定のオブジェクトを見たと判定された場合に、音声認識処理を制御するステップと、

を有する、情報処理装置により実行される情報処理方法。

(20)

表示画面におけるユーザの視線の位置に関する情報に基づいて、ユーザが所定のオブジェクトを見たかを判定するステップ、

ユーザが所定のオブジェクトを見たと判定された場合に、音声認識処理を制御するステップ、

をコンピュータに実行させるためのプログラム。

【符号の説明】

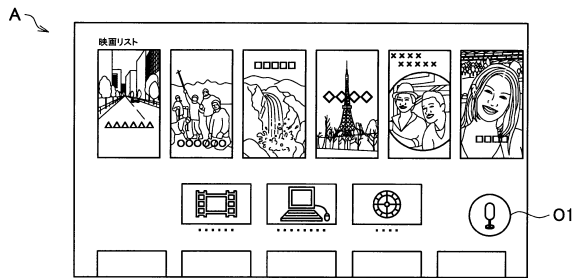
【0197】

- 100 情報処理装置
- 102 通信部
- 104 制御部
- 110 判定部
- 112 音声認識制御部
- 114 表示制御部

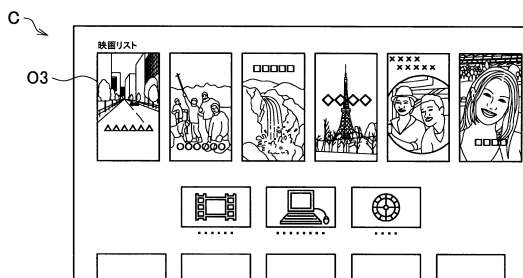
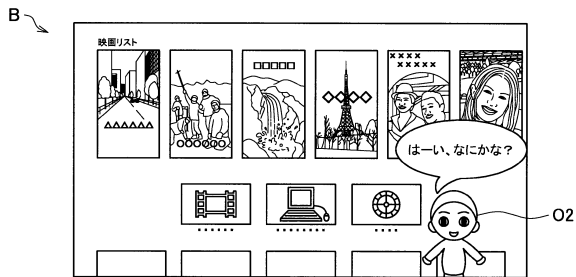
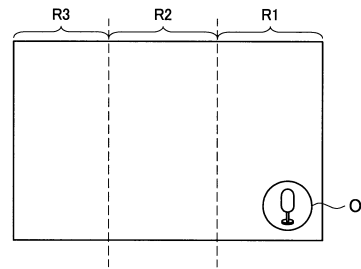
10

20

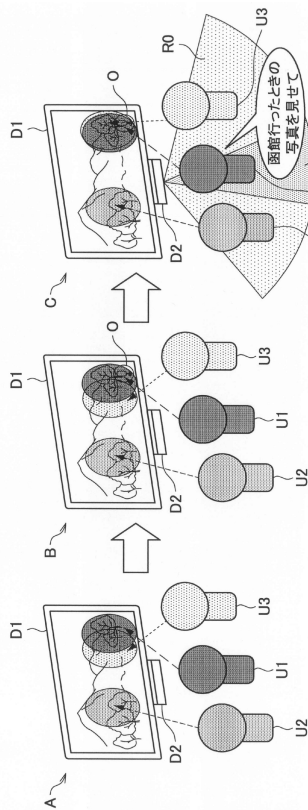
【図 1】



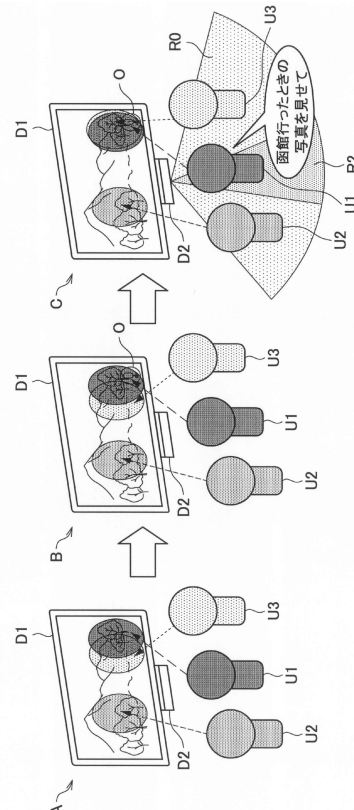
【図 2】



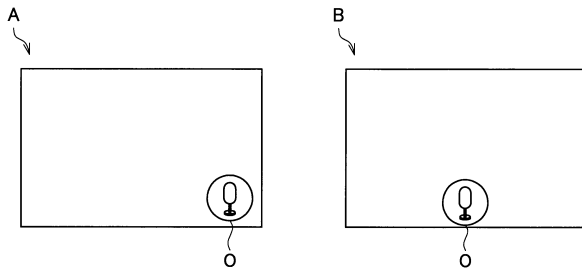
【図 3】



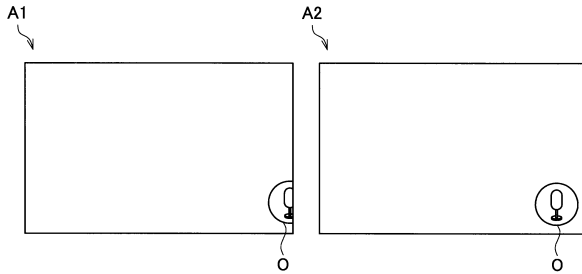
【図 4】



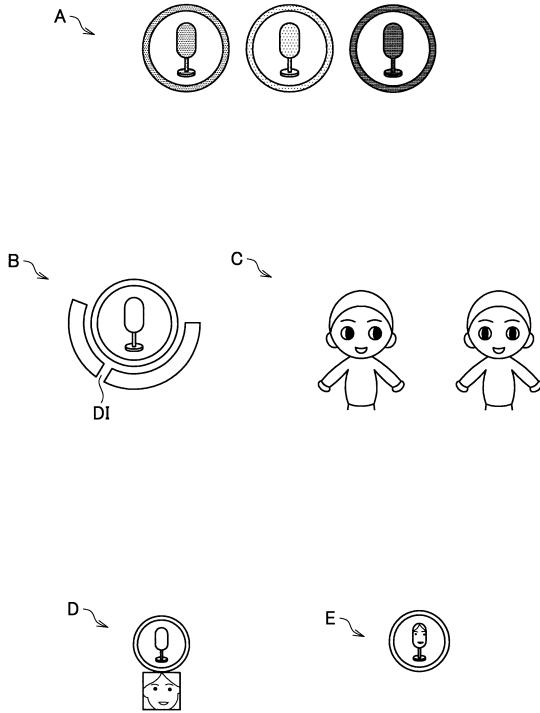
【図 5】



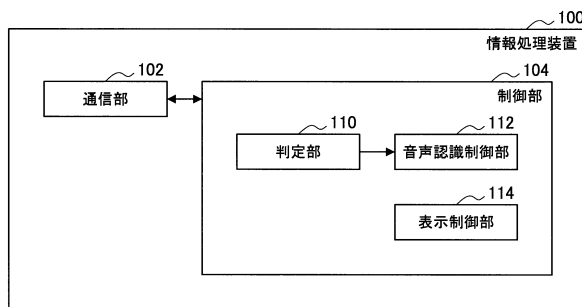
【図 6】



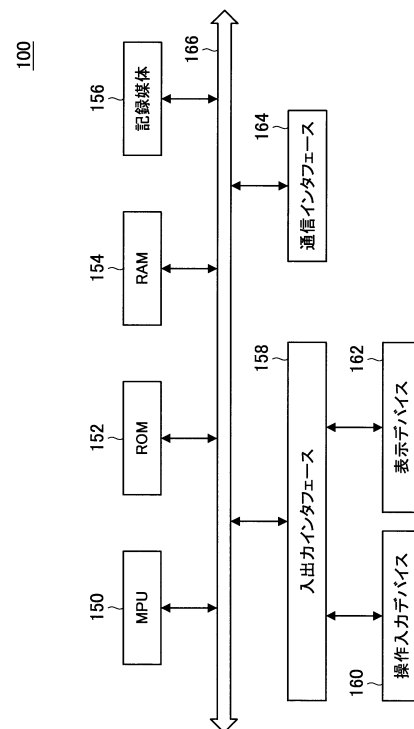
【図 7】



【図 8】



【図 9】



フロントページの続き

(51)Int.Cl. F I
G 0 6 F 3/16 6 3 0
G 0 6 F 3/16 6 5 0

(72)発明者 野田 卓郎
東京都港区港南 1 丁目 7 番 1 号 ソニー株式会社内
(72)発明者 安田 亮平
東京都港区港南 1 丁目 7 番 1 号 ソニー株式会社内

審査官 大野 弘

(56)参考文献 特開平 1 1 - 0 2 4 6 9 4 (J P , A)
特開平 1 0 - 2 6 0 7 7 3 (J P , A)
特開平 0 7 - 2 4 4 5 5 6 (J P , A)
特開 2 0 0 7 - 3 1 0 6 2 0 (J P , A)

(58)調査した分野(Int.Cl. , D B 名)
G 1 0 L 1 5 / 2 8
G 0 6 F 3 / 0 3 4 6
G 0 6 F 3 / 0 4 8 1
G 0 6 F 3 / 1 6
G 1 0 L 2 1 / 0 3 0 8