

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5772069号
(P5772069)

(45) 発行日 平成27年9月2日(2015.9.2)

(24) 登録日 平成27年7月10日(2015.7.10)

(51) Int.Cl. F I
HO4N 5/60 (2006.01) HO4N 5/60 C

請求項の数 11 (全 20 頁)

(21) 出願番号	特願2011-47892 (P2011-47892)	(73) 特許権者	000002185
(22) 出願日	平成23年3月4日(2011.3.4)		ソニー株式会社
(65) 公開番号	特開2012-186622 (P2012-186622A)		東京都港区港南1丁目7番1号
(43) 公開日	平成24年9月27日(2012.9.27)	(74) 代理人	100095957
審査請求日	平成26年1月21日(2014.1.21)		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	鶴見 辰吾
			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】

コンテンツの映像が表示される表示部の近傍に位置するユーザの画像を取得する画像取得部と、

前記画像に基づいて前記コンテンツに対する前記ユーザの視聴状態を判定する視聴状態判定部と、

前記視聴状態に応じて、前記ユーザに対する前記コンテンツの音声の出力を制御する音声出力制御部と、

前記コンテンツの各部分の重要度を判定する重要度判定部と

を備え、

前記音声出力制御部は、前記視聴状態として前記ユーザが前記音声を聴いていないことが判定された場合であって、前記重要度がより高い前記コンテンツの部分が出力されている場合に前記音声の音量を上げる情報処理装置。

【請求項2】

前記視聴状態判定部は、前記画像から検出される前記ユーザの目の開閉に基づいて、前記ユーザが前記音声を聴いているか否かを前記視聴状態として判定する、請求項1に記載の情報処理装置。

【請求項3】

前記視聴状態判定部は、前記画像から検出される前記ユーザの口の開閉に基づいて、前記ユーザが前記音声を聴いているか否かを前記視聴状態として判定する、請求項1に記載

の情報処理装置。

【請求項 4】

前記ユーザが発した音声を取得する音声取得部をさらに備え、

前記視聴状態判定部は、前記音声に含まれる発話の話者が前記ユーザであるか否かに基づいて、前記ユーザが前記音声を聴いているか否かを前記視聴状態として判定する、請求項 1 に記載の情報処理装置。

【請求項 5】

前記視聴状態判定部は、前記画像から検出される前記ユーザの向きに基づいて、前記ユーザが前記音声を聴いているか否かを前記視聴状態として判定する、請求項 1 に記載の情報処理装置。

10

【請求項 6】

前記視聴状態判定部は、前記画像から検出される前記ユーザの姿勢に基づいて、前記ユーザが前記音声を聴いているか否かを前記視聴状態として判定する、請求項 1 に記載の情報処理装置。

【請求項 7】

前記音声出力制御部は、前記視聴状態として前記ユーザが前記音声を聴いていないことが判定された場合に前記音声の音量を下げる、請求項 1 に記載の情報処理装置。

【請求項 8】

前記画像に含まれる顔によって前記ユーザを識別する顔識別部をさらに備え、

前記重要度判定部は、前記識別されたユーザの属性に基づいて前記重要度を判定する、請求項 1 に記載の情報処理装置。

20

【請求項 9】

前記画像に含まれる顔によって前記ユーザを識別する顔識別部をさらに備え、

前記視聴状態判定部は、前記画像に基づいて前記ユーザが前記コンテンツの映像を見ているか否かを判定し、

前記音声出力制御部は、前記識別されたユーザが前記映像を見ていると判定された場合に、前記識別されたユーザの属性に応じて前記音声の音質を変更する、請求項 1 に記載の情報処理装置。

【請求項 10】

コンテンツの映像が表示される表示部の近傍に位置するユーザの画像を取得することと

30

前記画像に基づいて前記コンテンツに対する前記ユーザの視聴状態を判定することと、
前記視聴状態に応じて、前記ユーザに対する前記コンテンツの音声の出力を制御することと、

前記コンテンツの各部分の重要度を判定することと

を含み、前記音声の出力を制御することは、前記視聴状態として前記ユーザが前記音声を聴いていないことが判定された場合であって、前記重要度がより高い前記コンテンツの部分が出力されている場合に前記音声の音量を上げることを含む情報処理方法。

【請求項 11】

コンテンツの映像が表示される表示部の近傍に位置するユーザの画像を取得する画像取得部と、

40

前記画像に基づいて前記コンテンツに対する前記ユーザの視聴状態を判定する視聴状態判定部と、

前記視聴状態に応じて、前記ユーザに対する前記コンテンツの音声の出力を制御する音声出力制御部と、

前記コンテンツの各部分の重要度を判定する重要度判定部と

としてコンピュータを動作させ、

前記音声出力制御部は、前記視聴状態として前記ユーザが前記音声を聴いていないことが判定された場合であって、前記重要度がより高い前記コンテンツの部分が出力されている場合に前記音声の音量を上げるプログラム。

50

【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】

TVなどの表示装置は、例えば住宅の居間、個室など至るところに設置され、生活のさまざまな局面でユーザにコンテンツの映像や音声を提供している。それゆえ、提供されるコンテンツに対するユーザの視聴状態も、さまざまである。ユーザは、必ずしも専らコンテンツを視聴するわけではなく、例えば、勉強や読書をしながらコンテンツを視聴したりする場合がある。そこで、コンテンツに対するユーザの視聴状態に合わせて、コンテンツの映像や音声の再生特性を制御する技術が開発されている。例えば、特許文献1には、ユーザの視線を検出することによってコンテンツに対するユーザの関心の程度を判定し、判定結果に応じてコンテンツの映像または音声の出力特性を変化させる技術が記載されている。

10

【先行技術文献】

【特許文献】

【0003】

【特許文献1】特開2004-312401号公報

【発明の概要】

20

【発明が解決しようとする課題】

【0004】

しかし、コンテンツに対するユーザの視聴状態はさらに多様化している。それゆえ、特許文献1に記載の技術では、それぞれの視聴状態におけるユーザの細かなニーズに対応したコンテンツの出力を提供するために十分ではない。

【0005】

そこで、視聴状態ごとのユーザのニーズにより的確に対応してコンテンツの出力を制御する技術が求められている。

【課題を解決するための手段】

【0006】

30

本開示によれば、コンテンツの映像が表示される表示部の近傍に位置するユーザの画像を取得する画像取得部と、上記画像に基づいて上記コンテンツに対する上記ユーザの視聴状態を判定する視聴状態判定部と、上記視聴状態に応じて、上記ユーザに対する上記音声の出力を制御する音声出力制御部と、上記コンテンツの各部分の重要度を判定する重要度判定部とを含み、上記音声出力制御部は、上記視聴状態として上記ユーザが上記音声を聴いていないことが判定された場合であって、上記重要度がより高い上記コンテンツの部分が出力されている場合に上記音声の音量を上げる情報処理装置が提供される。

【0007】

また、本開示によれば、コンテンツの映像が表示される表示部の近傍に位置するユーザの画像を取得することと、上記画像に基づいて上記コンテンツに対する上記ユーザの視聴状態を判定することと、上記視聴状態に応じて、上記ユーザに対する上記音声の出力を制御することと、上記コンテンツの各部分の重要度を判定することとを含み、上記音声の出力を制御することは、上記視聴状態として上記ユーザが上記音声を聴いていないことが判定された場合であって、上記重要度がより高い上記コンテンツの部分が出力されている場合に上記音声の音量を上げることを含む情報処理方法が提供される。

40

【0008】

また、本開示によれば、コンテンツの映像が表示される表示部の近傍に位置するユーザの画像を取得する画像取得部と、上記画像に基づいて上記コンテンツに対する上記ユーザの視聴状態を判定する視聴状態判定部と、上記視聴状態に応じて、上記ユーザに対する上記音声の出力を制御する音声出力制御部と、上記コンテンツの各部分の重要度を判定する

50

重要度判定部としてコンピュータを動作させ、上記音声出力制御部は、上記視聴状態として上記ユーザが上記音声を聴いていないことが判定された場合であって、上記重要度がより高い上記コンテンツの部分が出力されている場合に上記音声の音量を上げるプログラムが提供される。

【0009】

本開示によれば、例えば、コンテンツに対するユーザの視聴状態が、コンテンツの音声の出力制御に反映される。

【発明の効果】

【0010】

以上説明したように本開示によれば、視聴状態ごとのユーザのニーズによりの確に対応してコンテンツの出力を制御することができる。

10

【図面の簡単な説明】

【0011】

【図1】本開示の一実施形態に係る情報処理装置の機能構成を示すブロック図である。

【図2】本開示の一実施形態に係る情報処理装置の画像処理部の機能構成を示すブロック図である。

【図3】本開示の一実施形態に係る情報処理装置の音声処理部の機能構成を示すブロック図である。

【図4】本開示の一実施形態に係る情報処理装置のコンテンツ解析部の機能構成を示すブロック図である。

20

【図5】本開示の一実施形態における処理の例を示すフローチャートである

【図6】本開示の一実施形態に係る情報処理装置のハードウェア構成を説明するためのブロック図である。

【発明を実施するための形態】

【0012】

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、本明細書および図面において、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。

【0013】

なお、説明は以下の順序で行うものとする。

30

1. 機能構成
2. 処理フロー
3. ハードウェア構成
4. まとめ
5. 補足

【0014】

(1. 機能構成)

まず、図1を参照して、本開示の一実施形態に係る情報処理装置100の概略的な機能構成について説明する。図1は、情報処理装置100の機能構成を示すブロック図である。

40

【0015】

情報処理装置100は、画像取得部101、画像処理部103、音声取得部105、音声処理部107、視聴状態判定部109、音声出力制御部111、音声出力部113、コンテンツ取得部115、コンテンツ解析部117、重要度判定部119、およびコンテンツ情報記憶部151を含む。情報処理装置100は、例えば、TVチューナやPC(Personal Computer)などとして実現されうる。情報処理装置100には、表示装置10、カメラ20、およびマイク30に接続される。表示装置10は、コンテンツの映像が表示される表示部11と、コンテンツの音声出力されるスピーカ12とを含む。情報処理装置100は、これらの装置はと一体になったTV受像機やPCなどであってもよい。なお、表示装置10の表示部11にコンテンツの映像データを提供する構成など、コンテンツ再

50

生のための公知の構成が適用されうる部分については、図示を省略した。

【0016】

画像取得部101は、例えば、CPU(Central Processing Unit)、ROM(Read Only Memory)、RAM(Random Access Memory)、および通信装置などによって実現される。画像取得部101は、情報処理装置100に接続されたカメラ20から、表示装置10の表示部11の近傍に位置するユーザU1、U2の画像を取得する。なお、ユーザは、図示されているように複数であってもよく、また単一であってもよい。画像取得部101は、取得した画像の情報を画像処理部103に提供する。

【0017】

画像処理部103は、例えば、CPU、GPU(Graphics Processing Unit)、ROM、およびRAMなどによって実現される。画像処理部103は、画像取得部101から取得した画像の情報をフィルタリングなどによって処理し、ユーザU1、U2に関する情報を取得する。例えば、画像処理部103は、画像からユーザU1、U2の顔角度、口の開閉、目の開閉、視線方向、位置、姿勢などの情報を取得する。また、画像処理部103は、画像に含まれる顔の画像に基づいてユーザU1、U2を識別し、ユーザIDを取得してもよい。画像処理部103は、取得したこれらの情報を、視聴状態判定部109およびコンテンツ解析部117に提供する。なお、画像処理部103の詳細な機能構成については後述する。

10

【0018】

音声取得部105は、例えば、CPU、ROM、RAM、および通信装置などによって実現される。音声取得部105は、情報処理装置100に接続されたマイク30から、ユーザU1、U2が発した音声を取得する。音声取得部105は、取得した音声の情報を音声処理部107に提供する。

20

【0019】

音声処理部107は、例えば、CPU、ROM、およびRAMなどによって実現される。音声処理部107は、音声取得部105から取得した音声の情報をフィルタリングなどによって処理し、ユーザU1、U2が発した音声に関する情報を取得する。例えば、音声ユーザU1、U2の発話によるものである場合に、音声処理部107は、話者であるユーザU1、U2を推定してユーザIDを取得する。また、音声処理部107は、音声から音源方向、発話の有無などの情報を取得してもよい。音声処理部107は、取得したこれらの情報を、視聴状態判定部109に提供する。なお、音声処理部107の詳細な機能構成については後述する。

30

【0020】

視聴状態判定部109は、例えば、CPU、ROM、およびRAMなどによって実現される。視聴状態判定部109は、ユーザU1、U2の動作に基づいて、コンテンツに対するユーザU1、U2の視聴状態を判定する。ユーザU1、U2の動作は、画像処理部103、または音声処理部107から取得される情報に基づいて判定される。ユーザの動作は、例えば、「映像を見ている」、「目を瞑っている」、「口が会話の動きをしている」、「発話している」などである。このようなユーザの動作に基づいて判定されるユーザの視聴状態は、例えば、「通常視聴中」、「居眠り中」、「会話中」、「電話中」、「作業中」などである。視聴状態判定部109は、判定された視聴状態の情報を、音声出力制御部111に提供する。

40

【0021】

音声出力制御部111は、例えば、CPU、DSP(Digital Signal Processor)、ROM、およびRAMなどによって実現される。音声出力制御部111は、視聴状態判定部109から取得した視聴状態に応じて、ユーザに対するコンテンツの音声の出力を制御する。音声出力制御部111は、例えば、音声の音量を上げたり、音声の音量を下げたり、音声の音質を変更したりする。音声出力制御部111は、音声に含まれるボーカルの音量を上げるなど、音声の種類ごとに出力を制御してもよい。また、音声出力制御部111は、重要度判定部119から取得したコンテンツの各部分の重要度に応じて音声の出力を

50

制御してもよい。さらに、音声出力制御部 111 は、画像処理部 103 が取得したユーザ ID を用いて、ROM、RAM、およびストレージ装置などに予め登録されたユーザの属性情報を参照し、属性情報として登録されたユーザの好みに応じて音声の出力を制御してもよい。音声出力制御部 111 は、音声出力の制御情報を音声出力部 113 に提供する。

【0022】

音声出力部 113 は、例えば、CPU、DSP、ROM、および RAM などによって実現される。音声出力部 113 は、音声出力制御部 111 から取得した制御情報に従って、コンテンツの音声を表示装置 10 のスピーカ 12 に出力する。なお、出力の対象になるコンテンツの音声データは、図示しないコンテンツ再生のための構成によって音声出力部 113 に提供される。

10

【0023】

コンテンツ取得部 115 は、例えば、CPU、ROM、RAM、および通信装置などによって実現される。コンテンツ取得部 115 は、表示装置 10 によってユーザ U1、U2 に提供されるコンテンツを取得する。コンテンツ取得部 115 は、例えば、アンテナが受信した放送波を復調してデコードすることによって放送コンテンツを取得してもよい。また、コンテンツ取得部 115 は、通信装置を介して通信ネットワークからコンテンツをダウンロードしてもよい。さらに、コンテンツ取得部 115 は、ストレージ装置に格納されたコンテンツを読み出してもよい。コンテンツ取得部 115 は、取得したコンテンツの映像データおよび音声データを、コンテンツ解析部 117 に提供する。

【0024】

20

コンテンツ解析部 117 は、例えば、CPU、ROM、および RAM などによって実現される。コンテンツ解析部 117 は、コンテンツ取得部 115 から取得したコンテンツの映像データおよび音声のデータを解析して、コンテンツに含まれるキーワードや、コンテンツのシーンを検出する。コンテンツ取得部 115 は、画像処理部 103 から取得したユーザ ID を用いて、予め登録されたユーザの属性情報を参照し、ユーザ U1、U2 の関心が高いキーワードやシーンを検出する。コンテンツ解析部 117 は、これらの情報を重要度判定部 119 に提供する。なお、コンテンツ解析部 117 の詳細な機能構成については後述する。

【0025】

コンテンツ情報記憶部 151 は、例えば、ROM、RAM、およびストレージ装置などによって実現される。コンテンツ情報記憶部 151 には、例えば EPG、ECG などのコンテンツ情報が格納される。コンテンツ情報は、例えば、コンテンツ取得部 115 によってコンテンツとともに取得されてコンテンツ情報記憶部 151 に格納されてもよい。

30

【0026】

重要度判定部 119 は、例えば、CPU、ROM、および RAM などによって実現される。重要度判定部 119 は、コンテンツの各部分の重要度を判定する。重要度判定部 119 は、例えば、コンテンツ解析部 117 から取得したユーザの関心が高いキーワードやシーンの情報に基づいて、コンテンツの各部分の重要度を判定する。この場合、重要度判定部 119 は、かかるキーワードやシーンが検出されたコンテンツの部分を重要であると判定する。また、重要度判定部 119 は、コンテンツ情報記憶部 151 から取得されたコンテンツ情報に基づいてコンテンツの各部分の重要度を判定してもよい。この場合、重要度判定部 119 は、画像処理部 103 が取得したユーザ ID を用いて、予め登録されたユーザの属性情報を参照し、属性情報として登録されたユーザの好みに適合するコンテンツの部分を重要であると判定する。また、重要度判定部 119 は、コンテンツ情報によって示されるコマーシャルからコンテンツ本編への切り替わり部分など、ユーザに関わらず一般的に関心が高い部分を重要であると判定してもよい。

40

【0027】

(画像処理部の詳細)

続いて、図 2 を参照して、情報処理装置 100 の画像処理部 103 の機能構成についてさらに説明する。図 2 は、画像処理部 103 の機能構成を示すブロック図である。

50

【0028】

画像処理部103は、顔検出部1031、顔追跡部1033、顔識別部1035、および姿勢推定部1037を含む。顔識別部1035は、顔識別用DB153を参照する。画像処理部103は、画像取得部101から画像データを取得する。また、画像処理部103は、ユーザを識別するユーザID、および顔角度、口の開閉、目の開閉、視線方向、位置、姿勢などの情報を視聴状態判定部109またはコンテンツ解析部117に提供する。

【0029】

顔検出部1031は、例えば、CPU、GPU、ROM、およびRAMなどによって実現される。顔検出部1031は、画像取得部101から取得した画像データを参照して、画像に含まれる人間の顔を検出する。画像の中に顔が含まれている場合、顔検出部1031は、当該顔の位置や大きさなどを検出する。さらに、顔検出部1031は、画像によって示される顔の状態を検出する。例えば、顔検出部1031は、顔の角度、目を瞑っているか否か、視線の方向といったような状態を検出する。なお、顔検出部1031の処理には、例えば、特開2007-65766号公報や、特開2005-44330号公報に掲載されている技術など、公知のあらゆる技術を適用することが可能である。

【0030】

顔追跡部1033は、例えば、CPU、GPU、ROM、およびRAMなどによって実現される。顔追跡部1033は、画像取得部101から取得した異なるフレームの画像データについて、顔検出部1031によって検出された顔を追跡する。顔追跡部1033は、顔検出部1031によって検出された顔の画像データのパターンの類似性などを利用して、後続のフレームで当該顔に対応する部分を探索する。顔追跡部1033のこのような処理によって、複数のフレームの画像に含まれる顔が、同一のユーザの顔の時系列変化として認識されうる。

【0031】

顔識別部1035は、例えば、CPU、GPU、ROM、およびRAMなどによって実現される。顔識別部1035は、顔検出部1031によって検出された顔について、どのユーザの顔であるかの識別を行う処理部である。顔識別部1035は、顔検出部1031によって検出された顔の特徴的な部分などに着目して局所特徴量を算出し、算出した局所特徴量と、顔識別用DB153に予め格納されたユーザの顔画像の局所特徴量とを比較することによって、顔検出部1031により検出された顔を識別し、顔に対応するユーザのユーザIDを特定する。なお、顔識別部1035の処理には、例えば、特開2007-65766号公報や、特開2005-44330号公報に掲載されている技術など、公知のあらゆる技術を適用することが可能である。

【0032】

姿勢推定部1037は、例えば、CPU、GPU、ROM、およびRAMなどによって実現される。姿勢推定部1037は、画像取得部101から取得した画像データを参照して、画像に含まれるユーザの姿勢を推定する。姿勢推定部1037は、予め登録されたユーザの姿勢の種類ごとの画像の特徴などに基づいて、画像に含まれるユーザの姿勢がどのような種類の姿勢であるかを推定する。例えば、姿勢推定部1037は、ユーザが機器を保持して耳に近づけている姿勢が画像から認識される場合に、ユーザが電話中の姿勢であると推定する。なお、姿勢推定部1037の処理には、公知のあらゆる技術を適用することが可能である。

【0033】

顔識別用DB153は、例えば、ROM、RAM、およびストレージ装置などによって実現される。顔識別用DB153には、例えば、ユーザの顔画像の局所特徴量が、ユーザIDと関連付けて予め格納される。顔識別用DB153に格納されたユーザの顔画像の局所特徴量は、顔識別部1035によって参照される。

【0034】

(音声処理部の詳細)

続いて、図3を参照して、情報処理装置100の音声処理部107の機能構成について

10

20

30

40

50

さらに説明する。図3は、音声処理部107の機能構成を示すブロック図である。

【0035】

音声処理部107は、発話検出部1071、話者推定部1073、および音源方向推定部1075を含む。話者推定部1073は、話者識別用DB155を参照する。音声処理部107は、音声取得部105から音声データを取得する。また、音声処理部107は、ユーザを識別するユーザID、および音源方向、発話の有無などの情報を視聴状態判定部109に提供する。

【0036】

発話検出部1071は、例えば、CPU、ROM、およびRAMなどによって実現される。発話検出部1071は、音声取得部105から取得した音声データを参照して、音声に含まれる発話を検出する。音声の中に発話が含まれている場合、発話検出部1071は、当該発話の開始点、終了点、および周波数特性などを検出する。なお、発話検出部1071の処理には、公知のあらゆる技術を適用することが可能である。

10

【0037】

話者推定部1073は、例えば、CPU、ROM、およびRAMなどによって実現される。話者推定部1073は、発話検出部1071によって検出された発話について、話者を推定する。話者推定部1073は、例えば、発話検出部1071によって検出された発話の周波数特性などの特徴を、話者識別用DB155に予め登録されたユーザの発話音声の特徴と比較することによって、発話検出部1071によって検出された発話の話者を推定し、話者のユーザIDを特定する。なお、話者推定部1073の処理には、公知のあらゆる技術を適用することが可能である。

20

【0038】

音源方向推定部1075は、例えば、CPU、ROM、およびRAMなどによって実現される。音源方向推定部1075は、例えば、音声取得部105が位置の異なる複数のマイク30から取得した音声データの位相差を検出することによって、音声データに含まれる発話などの音声の音源の方向を推定する。音源方向推定部1075によって推定された音源の方向は、画像処理部103において検出されたユーザの位置と対応付けられ、これによって発話の話者が推定されてもよい。なお、音源方向推定部1075の処理には、公知のあらゆる技術を適用することが可能である。

【0039】

話者識別用DB155は、例えば、ROM、RAM、およびストレージ装置などによって実現される。話者識別用DB155には、例えば、ユーザの発話音声の周波数特性などの特徴が、ユーザIDと関連付けて予め格納される。話者識別用DB155に格納されたユーザの発話音声の特徴は、話者推定部1073によって参照される。

30

【0040】

(コンテンツ解析部の詳細)

続いて、図4を参照して、情報処理装置100のコンテンツ解析部117の機能構成についてさらに説明する。図4は、コンテンツ解析部117の機能構成を示すブロック図である。

【0041】

コンテンツ解析部117は、発話検出部1171、キーワード検出部1173、およびシーン検出部1175を含む。キーワード検出部1173は、キーワード検出用DB157を参照する。シーン検出部1175は、シーン検出用DB159を参照する。コンテンツ解析部117は、画像処理部103からユーザIDを取得する。また、コンテンツ解析部117は、コンテンツ取得部115からコンテンツの映像データおよび音声データを取得する。コンテンツ解析部117は、ユーザの関心が高いと推定されるキーワードやシーンの情報を重要度判定部119に提供する。

40

【0042】

発話検出部1171は、例えば、CPU、ROM、およびRAMなどによって実現される。発話検出部1171は、コンテンツ取得部115から取得したコンテンツの音声デー

50

タを参照して、音声に含まれる発話を検出する。音声の中に発話が含まれている場合、発話検出部 1171 は、当該発話の開始点、終了点、および周波数特性などの音声的特徴を検出する。なお、発話検出部 1171 の処理には、公知のあらゆる技術を適用することが可能である。

【0043】

キーワード検出部 1173 は、例えば、CPU、ROM、およびRAMなどによって実現される。キーワード検出部 1173 は、発話検出部 1171 によって検出された発話について、発話に含まれるキーワードを検出する。キーワードは、各ユーザの関心が高いキーワードとして予めキーワード検出用DB157に格納されている。キーワード検出部 1173 は、発話検出部 1171 によって検出された発話の区間から、キーワード検出用DB157に格納されているキーワードの音声的特徴を有する部分を探索する。キーワード検出部 1173 は、どのユーザの関心が高いキーワードを検出するかを決定するために、画像処理部 103 から取得したユーザIDを用いる。発話区間からキーワードが検出された場合、キーワード検出部 1173 は、例えば、検出されたキーワードと、当該キーワードへの関心が高いユーザのユーザIDとを関連づけて出力する。

10

【0044】

シーン検出部 1175 は、例えば、CPU、ROM、およびRAMなどによって実現される。シーン検出部 1175 は、コンテンツ取得部 115 から取得したコンテンツの映像データおよび音声データを参照して、コンテンツにおけるシーンを検出する。シーンは、各ユーザの関心が高いシーンとして予めシーン検出用DB159に格納されている。シーン検出部 1175 は、コンテンツの映像または音声が、シーン検出用DB159に格納されているシーンの映像的または音声的特徴を有するか否かを判定する。シーン検出部 1175 は、どのユーザの関心が高いシーンを検出するかを決定するために、画像処理部 103 から取得したユーザIDを用いる。シーンが検出された場合、シーン検出部 1175 は、例えば、検出されたシーンと、当該シーンへの関心が高いユーザのユーザIDとを関連付けて出力する。

20

【0045】

キーワード検出用DB157は、例えば、ROM、RAM、およびストレージ装置などによって実現される。キーワード検出用DB157には、例えば、ユーザの関心が高いキーワードの音声的特徴が、ユーザIDおよび当該キーワードを識別する情報と関連付けて予め格納される。キーワード検出用DB157に格納されたキーワードの音声的特徴は、キーワード検出部 1173 によって参照される。

30

【0046】

シーン検出用DB159は、例えば、ROM、RAM、およびストレージ装置などによって実現される。シーン検出用DB159には、例えば、ユーザの関心が高いシーンの映像的または音声的特徴が、ユーザIDおよび当該シーンを識別する情報と関連付けて予め格納される。シーン検出用DB159に格納されたシーンの映像的または音声的特徴は、シーン検出部 1175 によって参照される。

【0047】

(2. 処理フロー)

40

続いて、図5を参照して、本開示の一実施形態における処理フローについて説明する。図5は、本開示の一実施形態における視聴状態判定部109、音声出力制御部111、および重要度判定部119による処理の例を示すフローチャートである。

【0048】

図5を参照すると、まず、視聴状態判定部109が、ユーザU1、U2がコンテンツの映像を見ているか否かを判定する(ステップS101)。ここで、ユーザU1、U2が映像を見ているか否かは、画像処理部103において検出されるユーザU1、U2の顔角度、目の開閉、および視線方向によって判定されうる。例えば、視聴状態判定部109は、ユーザの顔角度および視線方向が表示装置10の表示部11の方向に近く、またユーザの目が瞑られていない場合に、「ユーザがコンテンツの映像を見ている」と判定する。ユー

50

ザU1, U2が複数である場合、視聴状態判定部109は、ユーザU1, U2のいずれかがコンテンツの映像を見ていると判定された場合に、「ユーザがコンテンツの映像を見ている」と判定しうる。

【0049】

ステップS101において、「ユーザがコンテンツの映像を見ている」と判定された場合、次に、視聴状態判定部109が、コンテンツに対するユーザの視聴状態は「通常視聴中」と判定する(ステップS103)。ここで、視聴状態判定部109は、視聴状態が「通常視聴中」であることを示す情報を音声出力制御部111に提供する。

【0050】

続いて、音声出力制御部111が、ユーザの好みに合わせて、コンテンツの音声の音質を変更する(ステップS105)。ここで、音声出力制御部111は、画像処理部103が取得したユーザIDを用いて、ROM、RAM、およびストレージ装置などに予め登録されたユーザの属性情報を参照し、属性情報として登録されたユーザの好みを取得しうる。

10

【0051】

一方、ステップS101において、「ユーザがコンテンツの映像を見ている」とは判定されなかった場合、次に、視聴状態判定部109が、ユーザU1, U2が目を瞑っているか否かを判定する(ステップS107)。ここで、ユーザU1, U2が目を瞑っているか否かは、画像処理部103において検出されるユーザU1, U2の目の開閉の時系列変化によって判定されうる。例えば、視聴状態判定部109は、ユーザの目が閉じた状態が所定の時間以上継続している場合に、「ユーザが目を瞑っている」と判定する。ユーザU1, U2が複数である場合、視聴状態判定部109は、ユーザU1, U2の両方が目を瞑っていると判定された場合に、「ユーザが目を瞑っている」と判定しうる。

20

【0052】

ステップS107において「ユーザが目を瞑っている」と判定された場合、次に、視聴状態判定部109が、コンテンツに対するユーザの視聴状態は「居眠り中」と判定する(ステップS109)。ここで、視聴状態判定部109は、視聴状態が「居眠り中」であることを示す情報を音声出力制御部111に提供する。

【0053】

続いて、音声出力制御部111が、コンテンツの音声の音量を徐々に小さくし、最終的に消音する(ステップS111)。かかる音声出力の制御によって、例えば、ユーザが居眠り中である場合にその居眠りを妨げないようにすることが可能である。このとき、音声出力の制御とともに、表示部11に表示される映像の輝度を下げ、最終的に消画する映像出力の制御が実行されてもよい。音量を徐々に小さくする途中でユーザの視聴状態が変わったり、ユーザから表示装置10への操作が取得されたりした場合、音量を小さくする制御は中止されうる。

30

【0054】

ここで、ステップS111における処理の変形例として、音声出力制御部111は、コンテンツの音声の音量を上げてよい。かかる音声出力の制御によって、例えば、ユーザがコンテンツを視聴したいにもかかわらず居眠りをしている場合にユーザをコンテンツの視聴に復帰させることが可能である。

40

【0055】

一方、ステップS107において、「ユーザが目を瞑っている」とは判定されなかった場合、次に、視聴状態判定部109が、ユーザU1, U2の口が会話中の動きになっているか否かを判定する(ステップS113)。ここで、ユーザU1, U2の口が会話中の動きになっているか否かは、画像処理部103において検出されるユーザU1, U2の口の開閉の時系列変化によって判定されうる。例えば、視聴状態判定部109は、ユーザの口の開閉が変化している状態が所定の時間以上継続している場合に、「ユーザの口が会話中の動きになっている」と判定する。ユーザU1, U2が複数である場合、視聴状態判定部109は、ユーザU1, U2のいずれかの口が会話中の動きになっている場合に、「ユー

50

ザの口が会話中の動きになっている」判定しうる。

【0056】

ステップS113において、「ユーザの口が会話中の動きになっている」と判定された場合、次に、視聴状態判定部109が、ユーザU1, U2の発話を検出されたか否かを判定する(ステップS115)。ここで、ユーザU1, U2の発話を検出されたか否かは、音声処理部107において検出される発話の話者のユーザIDによって判定されうる。例えば、視聴状態判定部109は、画像処理部103から取得したユーザIDが、音声処理部107から取得した発話の話者のユーザIDに一致する場合に、「ユーザの発話を検出された」と判定する。ユーザU1, U2が複数である場合、視聴状態判定部109は、ユーザU1, U2のいずれかの発話を検出された場合に、「ユーザの発話を検出された」と判定しうる。

10

【0057】

ステップS115において、「ユーザの発話を検出された」と判定された場合、次に、視聴状態判定部109が、ユーザU1, U2が別のユーザの方を向いているか否かを判定する(ステップS117)。ここで、ユーザU1, U2が別のユーザの方を向いているか否かは、画像処理部103において検出されるユーザU1, U2の顔角度、および位置によって判定されうる。例えば、視聴状態判定部109は、ユーザの顔角度によって示される当該ユーザが向いている方向が、他のユーザの位置と一致する場合に、「ユーザが別のユーザの方を向いている」と判定する。

【0058】

20

ステップS117において、「ユーザが別のユーザの方を向いている」と判定された場合、次に、視聴状態判定部109が、コンテンツに対するユーザの視聴状態は「会話中」と判定する(ステップS119)。ここで、視聴状態判定部109は、視聴状態が「会話中」であることを示す情報を音声出力制御部111に提供する。

【0059】

続いて、音声出力制御部111が、コンテンツの音声の音量をやや下げる(ステップS121)。かかる音声出力の制御によって、例えばユーザが会話中である場合にその会話を妨げないようにすることが可能になる。

【0060】

一方、ステップS117において「ユーザが別のユーザの方を向いている」とは判定されなかった場合、次に、視聴状態判定部109が、ユーザU1, U2が電話中の姿勢になっているか否かを判定する(ステップS123)。ここで、ユーザU1, U2が電話中の姿勢になっているか否かは、画像処理部103において検出されるユーザU1, U2の姿勢によって判定されうる。例えば、視聴状態判定部109は、画像処理部103に含まれる姿勢推定部1037が、ユーザが機器(受話器)を保持して耳に近づけている姿勢をユーザの電話中の姿勢であると推定した場合に、「ユーザが電話中の姿勢になっている」と判定する。

30

【0061】

ステップS123において「ユーザが電話中の姿勢になっている」と判定された場合、次に、視聴状態判定部109が、コンテンツに対するユーザの視聴状態は「電話中」と判定する(ステップS125)。ここで、視聴状態判定部109は、視聴状態が「電話中」であることを示す情報を音声出力制御部111に提供する。

40

【0062】

続いて、音声出力制御部111が、コンテンツの音声の音量をやや下げる(ステップS121)。かかる音声出力の制御によって、例えばユーザが電話中である場合にその電話を妨げないようにすることが可能になる。

【0063】

一方、ステップS113において「ユーザの口が会話中の動きになっている」とは判定されなかった場合、ステップS115において「ユーザの発話を検出された」とは判定されなかった場合、およびステップS123において「ユーザが電話中の姿勢になっている

50

」とは判定されなかった場合、次に、視聴状態判定部 109 が、コンテンツに対するユーザの視聴状態は「作業中」であると判定する（ステップ S127）。

【0064】

続いて、重要度判定部 119 が、ユーザ U1, U2 に提供中のコンテンツの重要度が高いか否かを判定する（ステップ S129）。ここで、提供中のコンテンツの重要度が高いか否かは、重要度判定部 119 において判定されるコンテンツの各部分の重要度によって判定される。例えば、重要度判定部 119 は、コンテンツ解析部 117 によってユーザの関心が高いキーワードやシーンが検出されたコンテンツの部分の重要度が高いと判定する。また、例えば、重要度判定部 119 は、コンテンツ情報記憶部 151 から取得されるコンテンツ情報によって、予め登録されたユーザの好みに適合するコンテンツの部分、またはコマーシャルからコンテンツ本編への切り替わり部分など一般的に関心が高い部分の重要度が高いと判定する。

10

【0065】

ステップ S129 において、コンテンツの重要度が高いと判定された場合、次に、音声出力制御部 111 が、コンテンツの音声のうち、ボーカルの音声の音量をやや上げる（ステップ S131）。かかる音声出力の制御によって、例えばユーザが表示装置 10 の近傍で読書、家事、勉強などコンテンツの視聴以外の作業をしている場合に、コンテンツの中でユーザの関心が高いと推定される部分が開始したことをユーザに知らせることが可能になる。

【0066】

20

（3. ハードウェア構成）

次に、図 6 を参照しながら、上記で説明された本開示の一実施形態に係る情報処理装置 100 のハードウェア構成について詳細に説明する。図 6 は、本開示の一実施形態に係る情報処理装置 100 のハードウェア構成を説明するためのブロック図である。

【0067】

情報処理装置 100 は、CPU901、ROM903、および RAM905 を含む。さらに、情報処理装置 100 は、ホストバス 907、ブリッジ 909、外部バス 911、インターフェース 913、入力装置 915、出力装置 917、ストレージ装置 919、ドライブ 921、接続ポート 923、および通信装置 925 を含んでもよい。

【0068】

30

CPU901 は、演算処理装置および制御装置として機能し、ROM903、RAM905、ストレージ装置 919、またはリムーバブル記録媒体 927 に記録された各種プログラムに従って、情報処理装置 900 内の動作全般またはその一部を制御する。ROM903 は、CPU901 が使用するプログラムや演算パラメータ等を記憶する。RAM905 は、CPU901 の実行において使用するプログラムや、その実行において適宜変化するパラメータ等を一次記憶する。これらは CPUバス等の内部バスにより構成されるホストバス 907 により相互に接続されている。

【0069】

ホストバス 907 は、ブリッジ 909 を介して、PCI (Peripheral Component Interconnect/Interface) バスなどの外部バス 911 に接続されている。

40

【0070】

入力装置 915 は、例えば、マウス、キーボード、タッチパネル、ボタン、スイッチおよびレバーなど、ユーザが操作する操作手段である。また、入力装置 915 は、例えば、赤外線やその他の電波を利用したリモートコントロール手段であってもよいし、情報処理装置 900 の操作に対応した携帯電話や PDA 等の外部接続機器 929 であってもよい。さらに、入力装置 915 は、例えば、上記の操作手段を用いてユーザにより入力された情報に基づいて入力信号を生成し、CPU901 に出力する入力制御回路などから構成されている。情報処理装置 900 のユーザは、この入力装置 915 を操作することにより、情報処理装置 900 に対して各種のデータを入力したり処理動作を指示したりすることができる。

50

【 0 0 7 1 】

出力装置 9 1 7 は、取得した情報をユーザに対して視覚的または聴覚的に通知することが可能な装置で構成される。このような装置として、C R Tディスプレイ装置、液晶ディスプレイ装置、プラズマディスプレイ装置、E Lディスプレイ装置およびランプなどの表示装置や、スピーカおよびヘッドホンなどの音声出力装置や、プリンタ装置、携帯電話、ファクシミリなどがある。出力装置 9 1 7 は、例えば、情報処置装置 9 0 0 が行った各種処理により得られた結果を出力する。具体的には、表示装置は、情報処置装置 9 0 0 が行った各種処理により得られた結果を、テキストまたはイメージで表示する。他方、音声出力装置は、再生された音声データや音響データ等からなるオーディオ信号をアナログ信号に変換して出力する。

10

【 0 0 7 2 】

ストレージ装置 9 1 9 は、情報処置装置 9 0 0 の記憶部の一例として構成されたデータ格納用の装置である。ストレージ装置 9 1 9 は、例えば、H D D (Hard Disk Drive) 等の磁気記憶部デバイス、半導体記憶デバイス、光記憶デバイス、または光磁気記憶デバイス等により構成される。このストレージ装置 9 1 9 は、C P U 9 0 1 が実行するプログラムや各種データ、および外部から取得した各種のデータなどを格納する。

【 0 0 7 3 】

ドライブ 9 2 1 は、記録媒体用リーダライタであり、情報処置装置 9 0 0 に内蔵、あるいは外付けされる。ドライブ 9 2 1 は、装着されている磁気ディスク、光ディスク、光磁気ディスク、または半導体メモリ等のリムーバブル記録媒体 9 2 7 に記録されている情報を読み出して、R A M 9 0 5 に出力する。また、ドライブ 9 2 1 は、装着されている磁気ディスク、光ディスク、光磁気ディスク、または半導体メモリ等のリムーバブル記録媒体 9 2 7 に記録を書き込むことも可能である。リムーバブル記録媒体 9 2 7 は、例えば、D V Dメディア、H D - D V Dメディア、B l u - r a y (登録商標)メディア等である。また、リムーバブル記録媒体 9 2 7 は、コンパクトフラッシュ (登録商標) (Compact Flash: C F)、フラッシュメモリ、または、S Dメモリカード (Secure Digital memory card) 等であってもよい。また、リムーバブル記録媒体 9 2 7 は、例えば、非接触型 I Cチップを搭載した I Cカード (Integrated Circuit card) または電子機器等であってもよい。

20

【 0 0 7 4 】

接続ポート 9 2 3 は、機器を情報処置装置 9 0 0 に直接接続するためのポートである。接続ポート 9 2 3 の一例として、U S B (Universal Serial Bus) ポート、I E E E 1 3 9 4 ポート、S C S I (Small Computer System Interface) ポート等がある。接続ポート 9 2 3 の別の例として、R S - 2 3 2 C ポート、光オーディオ端子、H D M I (High-Definition Multimedia Interface) ポート等がある。この接続ポート 9 2 3 に外部接続機器 9 2 9 を接続することで、情報処置装置 9 0 0 は、外部接続機器 9 2 9 から直接各種のデータを取得したり、外部接続機器 9 2 9 に各種のデータを提供したりする。

30

【 0 0 7 5 】

通信装置 9 2 5 は、例えば、通信ネットワーク 9 3 1 に接続するための通信デバイス等で構成された通信インターフェースである。通信装置 9 2 5 は、例えば、有線または無線 L A N (Local Area Network)、B l u e t o o t h (登録商標)、または W U S B (Wireless USB) 用の通信カード等である。また、通信装置 9 2 5 は、光通信用のルータ、A D S L (Asymmetric Digital Subscriber Line) 用のルータ、または、各種通信用のモデム等であってもよい。この通信装置 9 2 5 は、例えば、インターネットや他の通信機器との間で、例えば T C P / I P 等の所定のプロトコルに則して信号等を送受信することができる。また、通信装置 9 2 5 に接続される通信ネットワーク 9 3 1 は、有線または無線によって接続されたネットワーク等により構成され、例えば、インターネット、家庭内 L A N、赤外線通信、ラジオ波通信または衛星通信等であってもよい。

40

【 0 0 7 6 】

以上、情報処置装置 9 0 0 のハードウェア構成の一例を示した。上記の各構成要素は、

50

汎用的な部材を用いて構成されていてもよいし、各構成要素の機能に特化したハードウェアにより構成されていてもよい。従って、上記各実施形態を実施する時々の技術レベルに応じて、適宜、利用するハードウェア構成を変更することが可能である。

【0077】

(4.まとめ)

以上で説明された一実施形態によれば、コンテンツの映像が表示される表示部の近傍に位置するユーザの画像を取得する画像取得部と、画像に基づいてコンテンツに対するユーザの視聴状態を判定する視聴状態判定部と、視聴状態に応じて、ユーザに対する音声の出力を制御する音声出力制御部とを含む情報処理装置が提供される。

【0078】

この場合、例えば、ユーザがさまざまな事情でコンテンツの音声を聴いていない状態である場合を識別することによって、ユーザのニーズによりの確に対応してコンテンツの音声の出力を制御することができる。

【0079】

また、視聴状態判定部は、画像から検出されるユーザの目の開閉に基づいて、ユーザが音声を聴いているか否かを視聴状態として判定しうる。

【0080】

この場合、例えば、ユーザが居眠り中である場合などを識別して、コンテンツの音声の出力を制御することができる。例えばユーザが居眠り中である場合、コンテンツの音声に妨げられることなく居眠りをしたい、または居眠りを中止してコンテンツの視聴に復帰したいといったようなユーザのニーズが存在することが考えられる。上記の場合、このようなニーズによりの確に対応したコンテンツの音声の出力の制御が可能になる。

【0081】

また、視聴状態判定部は、画像から検出されるユーザの口の開閉に基づいて、ユーザが音声を聴いているか否かを視聴状態として判定しうる。

【0082】

この場合、例えば、ユーザが会話中、または電話中である場合などを識別して、コンテンツの音声の出力を制御することができる。例えばユーザが会話中または電話中である場合、コンテンツの音声が発話または電話の妨げになるために音量を小さくしたいといったようなユーザのニーズが存在することが考えられる。上記の場合、このようなニーズによりの確に対応したコンテンツの音声の出力の制御が可能になる。

【0083】

また、情報処理装置は、ユーザが発した音声を取得する音声取得部をさらに含み、視聴状態判定部は、音声に含まれる発話の話者がユーザであるか否かに基づいて、ユーザが音声を聴いているか否かを視聴状態として判定しうる。

【0084】

この場合、例えば、ユーザの口は開閉しているが発話はしていないような場合に、ユーザが会話中または電話中であると誤判定することを防ぐことができる。

【0085】

また、視聴状態判定部は、画像から検出されるユーザの向きに基づいて、ユーザが音声を聴いているか否かを視聴状態として判定しうる。

【0086】

この場合、例えば、ユーザが独り言を言っているような場合に、ユーザが会話中であると誤判定することを防ぐことができる。

【0087】

また、視聴状態判定部は、画像から検出されるユーザの姿勢に基づいて、ユーザが音声を聴いているか否かを視聴状態として判定しうる。

【0088】

この場合、例えば、ユーザが独り言を言っているような場合に、ユーザが電話中であると誤判定することを防ぐことができる。

10

20

30

40

50

【 0 0 8 9 】

また、音声出力制御部は、視聴状態としてユーザが音声を聴いていないことが判定された場合に音声の音量を下げてよい。

【 0 0 9 0 】

この場合、例えば、ユーザが居眠り中、会話中、または電話中などでコンテンツの音声を聴いておらず、それゆえコンテンツの音声を必要としていない場合、およびコンテンツの音声が邪魔になる場合などに、ユーザのニーズを反映してコンテンツの音声出力を制御することができる。

【 0 0 9 1 】

また、音声出力制御部は、視聴状態としてユーザが音声を聴いていないことが判定された場合に音声の音量を上げてよい。

10

【 0 0 9 2 】

この場合、例えば、ユーザが居眠り中、または作業中などでコンテンツの音声を聴いておらず、しかし、コンテンツの視聴に復帰することを望んでいるような場合に、ユーザのニーズを反映してコンテンツの音声出力を制御することができる。

【 0 0 9 3 】

また、情報処理装置は、コンテンツの各部分の重要度を判定する重要度判定部をさらに含み、音声出力制御部は、重要度がより高いコンテンツの部分で音声の音量を上げてよい。

【 0 0 9 4 】

20

この場合、例えば、ユーザが、コンテンツの特に重要な部分に限って、コンテンツの視聴に復帰することを望んでいるような場合に、ユーザのニーズを反映してコンテンツの音声出力を制御することができる。

【 0 0 9 5 】

また、情報処理装置は、画像に含まれる顔によってユーザを識別する顔識別部をさらに含み、重要度判定部は、識別されたユーザの属性に基づいて重要度を判定しうる。

【 0 0 9 6 】

この場合、例えば、画像によって自動的にユーザを識別し、さらに、識別されたユーザの好みを反映してコンテンツの重要部分を決定することができる。

【 0 0 9 7 】

30

また、情報処理装置は、画像に含まれる顔によってユーザを識別する顔識別部をさらに含み、視聴状態判定部は、画像に基づいてユーザがコンテンツの映像を見ているか否かを判定し、音声出力制御部は、識別されたユーザが映像を見ていると判定された場合に、識別されたユーザの属性に応じて音声の音質を変更しうる。

【 0 0 9 8 】

この場合、例えば、ユーザがコンテンツを視聴している場合に、ユーザの好みに合わせたコンテンツの音声出力を提供することができる。

【 0 0 9 9 】

(5 . 補 足)

上記実施形態では、ユーザの動作として「映像を見ている」、「目を瞑っている」、「口が会話の動きをしている」、「発話している」などを例示し、ユーザの視聴状態として「通常視聴中」、「居眠り中」、「会話中」、「電話中」、「作業中」などを例示したが、本技術はかかる例に限定されない。取得された画像および音声に基づいて、さまざまなユーザの動作および視聴状態が判定されうる。

40

【 0 1 0 0 】

また、上記実施形態では、ユーザの画像と、ユーザが発した音声に基づいてユーザの視聴状態を判定することとしたが、本技術はかかる例に限定されない。ユーザが発した音声は必ずしも視聴状態の判定に用いられなくてもよく、専らユーザの画像に基づいて視聴状態が判定されてもよい。

【 0 1 0 1 】

50

なお、本技術は以下のような構成も取ることができる。

(1) コンテンツの映像が表示される表示部の近傍に位置するユーザの画像を取得する画像取得部と、

前記画像に基づいて前記コンテンツに対する前記ユーザの視聴状態を判定する視聴状態判定部と、

前記視聴状態に応じて、前記ユーザに対する前記コンテンツの音声の出力を制御する音声出力制御部と、

を備える情報処理装置。

(2) 前記視聴状態判定部は、前記画像から検出される前記ユーザの目の開閉に基づいて、前記ユーザが前記音声を聴いているか否かを前記視聴状態として判定する、前記(1)に記載の情報処理装置。

10

(3) 前記視聴状態判定部は、前記画像から検出される前記ユーザの口の開閉に基づいて、前記ユーザが前記音声を聴いているか否かを前記視聴状態として判定する、前記(1)または(2)に記載の情報処理装置。

(4) 前記ユーザが発した音声を取得する音声取得部をさらに備え、前記視聴状態判定部は、前記音声に含まれる発話の話者が前記ユーザであるか否かに基づいて、前記ユーザが前記音声を聴いているか否かを前記視聴状態として判定する、前記(1)~(3)のいずれか1項に記載の情報処理装置。

(5) 前記視聴状態判定部は、前記画像から検出される前記ユーザの向きに基づいて、前記ユーザが前記音声を聴いているか否かを前記視聴状態として判定する、前記(1)~(4)のいずれか1項に記載の情報処理装置。

20

(6) 前記視聴状態判定部は、前記画像から検出される前記ユーザの姿勢に基づいて、前記ユーザが前記音声を聴いているか否かを前記視聴状態として判定する、前記(1)~(5)のいずれか1項に記載の情報処理装置。

(7) 前記音声出力制御部は、前記視聴状態として前記ユーザが前記音声を聴いていないことが判定された場合に前記音声の音量を下げる、前記(1)~(6)のいずれか1項に記載の情報処理装置。

(8) 前記音声出力制御部は、前記視聴状態として前記ユーザが前記音声を聴いていないことが判定された場合に前記音声の音量を上げる、前記(1)~(6)のいずれか1項に記載の情報処理装置。

30

(9) 前記コンテンツの各部分の重要度を判定する重要度判定部をさらに備え、前記音声出力制御部は、前記重要度がより高い前記コンテンツの部分で前記音声の音量を上げる、前記(8)に記載の情報処理装置。

(10) 前記画像に含まれる顔によって前記ユーザを識別する顔識別部をさらに備え、前記重要度判定部は、前記識別されたユーザの属性に基づいて前記重要度を判定する、前記(9)に記載の情報処理装置。

(11) 前記画像に含まれる顔によって前記ユーザを識別する顔識別部をさらに備え、前記視聴状態判定部は、前記画像に基づいて前記ユーザが前記コンテンツの映像を見ているか否かを判定し、

前記音声出力制御部は、前記識別されたユーザが前記映像を見ていると判定された場合に、前記識別されたユーザの属性に応じて前記音声の音質を変更する、前記(1)~(10)のいずれか1項に記載の情報処理装置。

40

(12) コンテンツの映像が表示される表示部の近傍に位置するユーザの画像を取得することと、

前記画像に基づいて前記コンテンツに対する前記ユーザの視聴状態を判定することと、前記視聴状態に応じて、前記ユーザに対する前記コンテンツの音声の出力を制御することと、

を含む情報処理方法。

(13) コンテンツの映像が表示される表示部の近傍に位置するユーザの画像を取得する画像取得部と、

50

前記画像に基づいて前記コンテンツに対する前記ユーザの視聴状態を判定する視聴状態判定部と、

前記視聴状態に応じて、前記ユーザに対する前記コンテンツの音声の出力を制御する音声出力制御部と、

としてコンピュータを動作させるプログラム。

【0102】

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、本技術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。

10

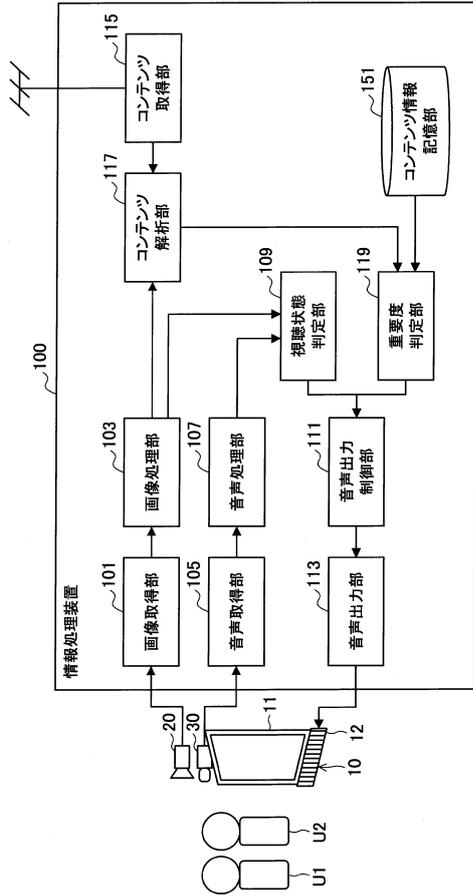
【符号の説明】

【0103】

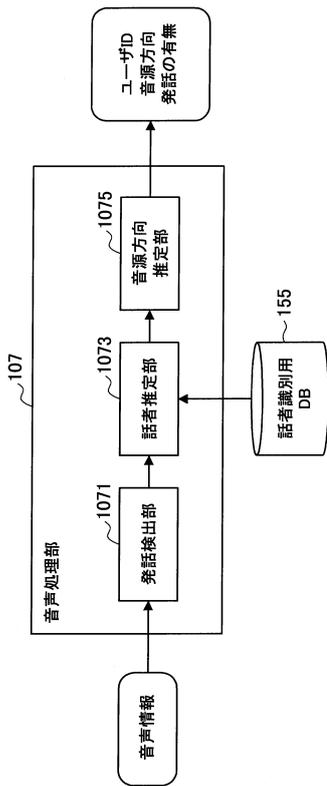
U 1 , U 2	ユーザ
1 0	表示装置
1 1	表示部
1 2	スピーカ
2 0	カメラ
3 0	マイク
1 0 0	情報処理装置
1 0 1	画像取得部
1 0 3	画像処理部
1 0 3 5	顔識別部
1 0 5	音声取得部
1 0 9	視聴状態判定部
1 1 1	音声出力制御部
1 1 3	音声出力部
1 1 9	重要度判定部

20

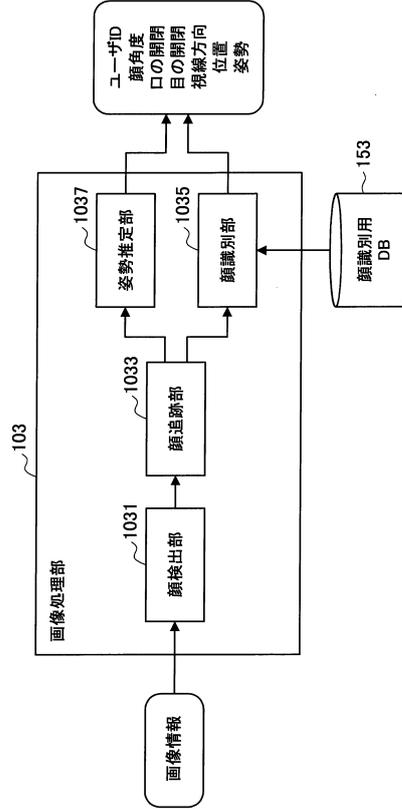
【図1】



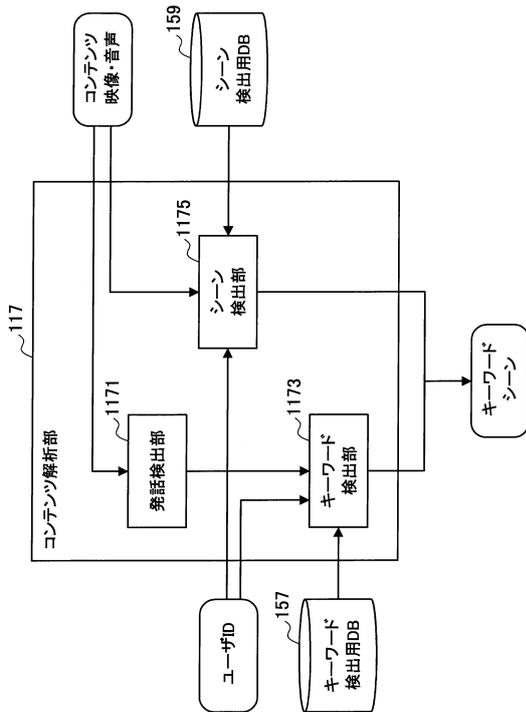
【図3】



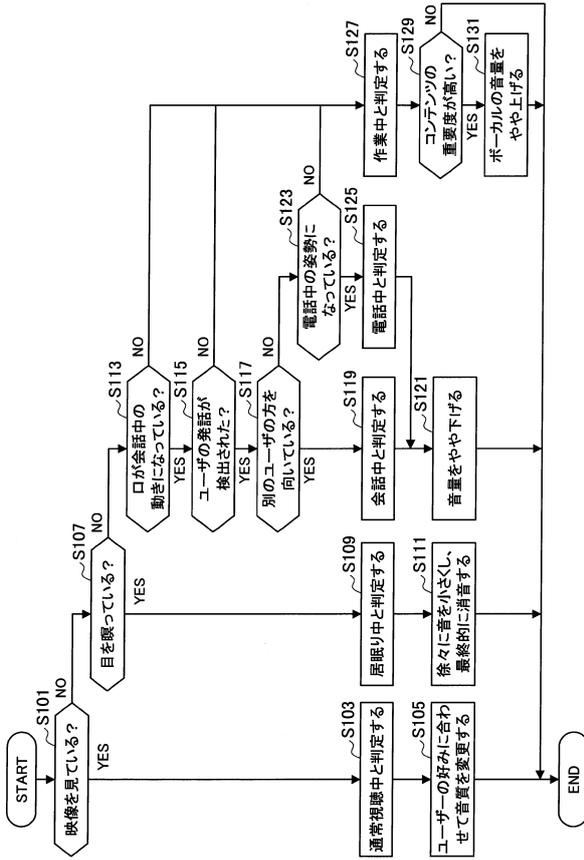
【図2】



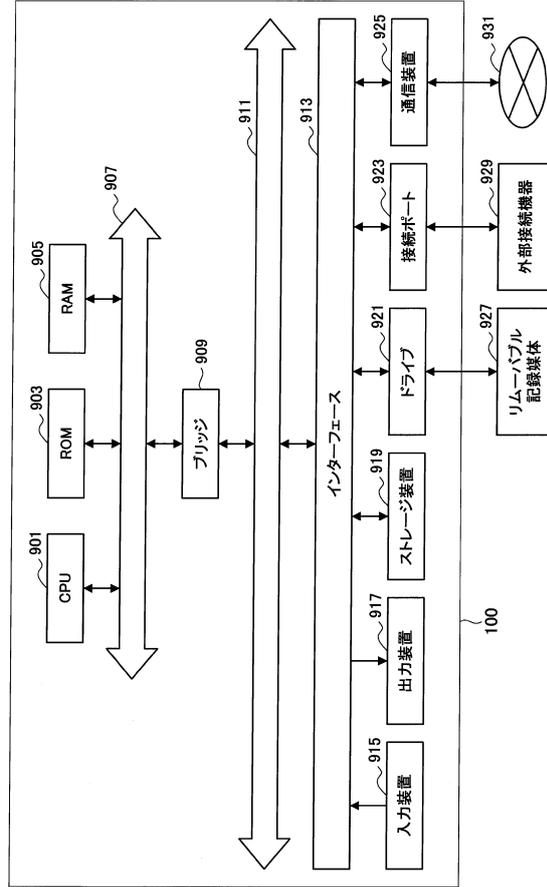
【図4】



【 図 5 】



【 図 6 】



フロントページの続き

審査官 大室 秀明

- (56)参考文献 特開平08 - 009282 (JP, A)
特開平09 - 034424 (JP, A)
特開2010 - 023639 (JP, A)
国際公開第2009 / 101622 (WO, A1)
特開2002 - 311977 (JP, A)
特開2006 - 005418 (JP, A)
特開2000 - 196970 (JP, A)

(58)調査した分野(Int.Cl., DB名)

H04N 5 / 38 - 5 / 63

H04R 3 / 00 - 3 / 14