

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B1)

(11) 特許番号

特許第6641045号
(P6641045)

(45) 発行日 令和2年2月5日(2020.2.5)

(24) 登録日 令和2年1月7日(2020.1.7)

(51) Int.Cl. F 1
G 0 6 T 1 3 / 2 0 (2 0 1 1 . 0 1) G 0 6 T 1 3 / 2 0 5 0 0

請求項の数 13 (全 16 頁)

<p>(21) 出願番号 特願2019-33926 (P2019-33926) (22) 出願日 平成31年2月27日 (2019.2.27) 審査請求日 平成31年3月20日 (2019.3.20) 早期審査対象出願</p>	<p>(73) 特許権者 518301442 みんとる合同会社 神奈川県横浜市西区高島2丁目11-2 スカイメナー横浜422 (74) 代理人 110000198 特許業務法人湘洋内外特許事務所 (72) 発明者 鈴木 智也 神奈川県横浜市西区高島2丁目11-2 スカイメナー横浜422 みんとる合同会社 社内 審査官 村松 貴士</p>
--	--

最終頁に続く

(54) 【発明の名称】 コンテンツ生成システム、及びコンテンツ生成方法

(57) 【特許請求の範囲】

【請求項1】

発話者の音声情報を取得する音声取得部と、
前記音声情報をテキスト情報に変換する変換部と、
前記テキスト情報を解析する解析部と、
前記テキスト情報の解析結果に基づいてシナリオ情報を生成するシナリオ情報生成部と

、
素材画像が格納された素材画像データベースと、
前記シナリオ情報を所定の編集単位に区分し、前記編集単位毎に前記素材画像データベースに格納された前記素材画像を対応付けて対応情報を生成する対応付け部と、
前記編集単位毎に対応付けられた前記素材画像を連結して動画像を生成する動画像生成部と、

を備え、

前記解析部は、前記音声情報に基づいて前記発話者のメタデータを特定し、

前記対応付け部は、前記発話者の前記メタデータに基づき、前記編集単位毎に前記素材画像データベースに格納された前記素材画像を対応付けて前記対応情報を生成する

ことを特徴とするコンテンツ生成システム。

【請求項2】

請求項1に記載のコンテンツ生成システムであって、

前記解析部は、前記テキスト情報の前記解析結果として、コンテンツにおける5W1H

を特定する

ことを特徴とするコンテンツ生成システム。

【請求項 3】

請求項 1 または 2 に記載のコンテンツ生成システムであって、
前記解析部は、コンテンツにおける登場人物の感情を特定する
ことを特徴とするコンテンツ生成システム。

【請求項 4】

請求項 3 に記載のコンテンツ生成システムであって、
前記解析部は、前記音声情報に基づいて前記コンテンツにおける登場人物の感情を特定
する

10

ことを特徴とするコンテンツ生成システム。

【請求項 5】

請求項 1 ~ 4 のいずれか一項に記載のコンテンツ生成システムであって、
前記解析部は、前記音声情報に基づいて前記発話者のメタデータとして、年代、性別、
出身国、使用言語、出身地方、及び使用言語の訛のうちの少なくとも一つを特定する
ことを特徴とするコンテンツ生成システム。

【請求項 6】

請求項 1 ~ 5 のいずれか一項に記載のコンテンツ生成システムであって、
前記対応付け部は、前記発話者の前記メタデータに基づき、前記編集単位毎に前記素材
画像データベースに格納された、所定の作風の前記素材画像を対応付けて前記対応情報を
生成する

20

ことを特徴とするコンテンツ生成システム。

【請求項 7】

請求項 6 に記載のコンテンツ生成システムであって、
前記素材画像データベースには、前記作風として、万人向け、男性向け、女性向け、子
供向け、成人向け、時代劇、西洋風、東洋風、実写、及びアニメーションのうちの少なく
とも一つの前記素材画像が格納されている
ことを特徴とするコンテンツ生成システム。

【請求項 8】

請求項 1 ~ 7 のいずれか一項に記載のコンテンツ生成システムであって、
前記シナリオ情報生成部は、前記テキスト情報の解析結果に基づいて、コンテンツにお
ける登場人物に関する少なくとも発言、挙動、状況、及び感情のうちの一つを時系列に配
置した前記シナリオ情報を生成する

30

ことを特徴とするコンテンツ生成システム。

【請求項 9】

請求項 1 ~ 8 のいずれか一項に記載のコンテンツ生成システムであって、
ユーザからの操作入力を受け付ける操作部と、
前記ユーザからの操作入力に基づき、前記シナリオ情報及び前記対応情報の少なくとも
一方を修正する修正部と、

を備えることを特徴とするコンテンツ生成システム。

40

【請求項 10】

請求項 9 に記載のコンテンツ生成システムであって、
前記修正部は、前記シナリオ情報及び前記対応情報の少なくとも一方に対する修正結果
を学習し、学習結果に基づいて前記素材画像データベースを更新する
ことを特徴とするコンテンツ生成システム。

【請求項 11】

請求項 9 に記載のコンテンツ生成システムであって、
前記修正部は、前記ユーザから入力された前記素材画像を前記シナリオ情報の前記編集
単位に対応付けることによって前記対応情報を修正する

ことを特徴とするコンテンツ生成システム。

50

【請求項 1 2】

請求項 1 ~ 1 1 のいずれか一項に記載のコンテンツ生成システムであって、
フリー素材としての画像データを収集し、前記素材画像として前記素材画像データベースに登録する収集部を、
備えることを特徴とするコンテンツ生成システム。

【請求項 1 3】

コンテンツ生成システムによるコンテンツ生成方法であって、
発話者の音声情報を取得する音声取得ステップと、
前記音声情報をテキスト情報に変換する変換ステップと、
前記テキスト情報を解析する解析ステップと、
前記テキスト情報の解析結果に基づいてシナリオ情報を生成するシナリオ情報生成ステップと、
前記シナリオ情報を所定の編集単位に区分し、前記編集単位毎に素材画像データベースに格納された素材画像を対応付けて対応情報を生成する対応付けステップと、
前記編集単位毎に対応付けられた前記素材画像を連結して動画像を生成する動画生成ステップと、
を含み、

前記解析ステップは、前記音声情報に基づいて前記発話者のメタデータを特定し、
前記対応付けステップは、前記発話者の前記メタデータに基づき、前記編集単位毎に前記素材画像データベースに格納された前記素材画像を対応付けて前記対応情報を生成することを特徴とするコンテンツ生成方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、例えば、発話音声に基づいてコンテンツを生成できるようにしたコンテンツ生成システム、及びコンテンツ生成方法に関する。

【背景技術】

【0002】

従来、動画像（アニメーションを含む）、漫画等のコンテンツを容易に生成するための様々方法が提案されている。

【0003】

例えば、特許文献 1 には、任意の動画データが指定されると動画データに含まれる複数の画像の中から重要なシーンの画像を自動的に抽出し、抽出した画像を漫画的に配置変換して表示し、画像には吹き出しなどの画面効果等を自動的に配置する、こと等が記載されている。

【0004】

また例えば、特許文献 2 には、入力された映像信号及びそれに対応する音声信号に基づき、映像信号から映像内容の切り替わりの特徴となる映像フレームを、漫画の構成単位であるコマ画像として抽出し、前記コマ画像に登場する人物の人物領域を検出し、前記音声信号から音声認識された前記人物の台詞を文字列情報として生成し、前記文字列情報を前記人物の台詞内容として挿入した吹き出しを、前記人物領域に基づいて前記コマ画像に重畳する、こと等が記載されている。

【先行技術文献】

【特許文献】

【0005】

【特許文献 1】特開 2 0 1 4 - 6 9 1 2 号公報

【特許文献 2】特開 2 0 0 3 - 8 5 5 7 2 号公報

【発明の概要】

【発明が解決しようとする課題】

【0006】

10

20

30

40

50

上述したように、動画像から画像を抽出して漫画を生成する技術は存在する。しかしながら、発話音声に基づいて動画像、漫画等のコンテンツを生成することはできない。

【0007】

本発明は、このような状況に鑑みてなされたものであり、発話音声に基づいてコンテンツを生成できるようにすることを目的とする。

【課題を解決するための手段】

【0008】

本発明は、上記課題の少なくとも一部を解決する手段を複数含んでいるが、その例を挙げるならば、以下のとおりである。

【0009】

上記課題を解決すべく、本発明の一態様に係るコンテンツ生成システムは、発話者の音声情報を取得する音声取得部と、前記音声情報をテキスト情報に変換する変換部と、前記テキスト情報を解析する解析部と、前記テキスト情報の解析結果に基づいてシナリオ情報を生成するシナリオ情報生成部と、素材画像が格納された素材画像データベースと、前記シナリオ情報を所定の編集単位に区分し、前記編集単位毎に前記素材画像データベースに格納された前記素材画像を対応付けて対応情報を生成する対応付け部と、前記編集単位毎に対応付けられた前記素材画像を連結して動画像を生成する動画像生成部と、を備えることを特徴とする。

【0010】

前記解析部は、前記テキスト情報の前記解析結果として、コンテンツにおける5W1Hを特定することができる。

【0011】

前記解析部は、コンテンツにおける登場人物の感情を特定することができる。

【0012】

前記解析部は、前記音声情報に基づいて前記コンテンツにおける登場人物の感情を特定することができる。

【0013】

前記解析部は、前記音声情報に基づいて前記発話者のメタデータを特定することができる。

【0014】

前記対応付け部は、前記発話者の前記メタデータに基づき、前記編集単位毎に前記素材画像データベースに格納された前記素材画像を対応付けて前記対応情報を生成することができる。

【0015】

前記対応付け部は、前記編集単位毎に前記素材画像データベースに格納された、所定の作風の前記素材画像を対応付けて前記対応情報を生成することができる。

【0016】

前記シナリオ情報生成部は、前記テキスト情報の解析結果に基づいて、コンテンツにおける登場人物に関する少なくとも発言、挙動、状況、及び感情のうちの一つを時系列に配置した前記シナリオ情報を生成することができる。

【0017】

前記コンテンツ生成システムは、ユーザからの操作入力を受け付ける操作部と、前記ユーザからの操作入力に基づき、前記シナリオ情報及び前記対応情報の少なくとも一方を修正する修正部と、を備えることができる。

【0018】

前記修正部は、前記シナリオ情報及び前記対応情報の少なくとも一方に対する修正結果を学習し、学習結果に基づいて前記素材画像データベースを更新することができる。

【0019】

前記修正部は、前記ユーザから入力された前記素材画像を前記シナリオ情報の前記編集単位に対応付けることによって前記対応情報を修正することができる。

10

20

30

40

50

【 0 0 2 0 】

前記コンテンツ生成システムは、フリー素材としての画像データを収集し、前記素材画像として前記素材画像データベースに登録する収集部を、備えることができる。

【 0 0 2 1 】

本発明の他の態様に係るコンテンツ生成方法は、発話者の音声情報を取得する音声取得ステップと、前記音声情報をテキスト情報に変換する変換ステップと、前記テキスト情報を解析する解析ステップと、前記テキスト情報の解析結果に基づいてシナリオ情報を生成するシナリオ情報生成ステップと、前記シナリオ情報を所定の編集単位に区分し、前記編集単位毎に素材画像データベースに格納された素材画像を対応付けて対応情報を生成する対応付けステップと、前記編集単位毎に対応付けられた前記素材画像を連結して動画画像を生成する動画生成ステップと、を含むことを特徴とする。

10

【発明の効果】

【 0 0 2 2 】

本発明の一態様によれば、発話音声に基づいてコンテンツを生成することが可能となる。

【 0 0 2 3 】

上記した以外の課題、構成及び効果は、以下の実施形態の説明により明らかにされる。

【図面の簡単な説明】

【 0 0 2 4 】

【図 1】図 1 は、本発明の一実施の形態に係るコンテンツ生成システムの構成例を示す図である。

20

【図 2】図 2 は、端末装置を構成する機能ブロックの構成例を示すブロック図である。

【図 3】図 3 は、サーバ装置を構成する機能ブロックの構成例を示すブロック図である。

【図 4】図 4 は、シナリオ情報のデータ構造の一例を示す図である。

【図 5】図 5 は、素材画像 DB のデータ構造の一例を示す図である。

【図 6】図 6 は、対応情報のデータ構造の一例を示す図である。

【図 7】図 7 は、コンテンツ生成処理の一例を説明するフローチャートである。

【図 8】図 8 は、修正処理の一例を説明するフローチャートである。

【図 9】図 9 は、コンテンツ表示画面の表示例を示す図である。

【図 10】図 10 は、コンピュータの構成例を示すブロック図である。

30

【発明を実施するための形態】

【 0 0 2 5 】

以下、本発明に係る一実施の形態を図面に基づいて説明する。なお、実施の形態を説明するための全図において、同一の部材には原則として同一の符号を付し、その繰り返しの説明は省略する。また、以下の実施の形態において、その構成要素（要素ステップ等も含む）は、特に明示した場合および原理的に明らかに必須であると考えられる場合等を除き、必ずしも必須のものではないことは言うまでもない。また、「A からなる」、「A よりなる」、「A を有する」、「A を含む」と言うときは、特にその要素のみである旨明示した場合等を除き、それ以外の要素を排除するものでないことは言うまでもない。同様に、以下の実施の形態において、構成要素等の形状、位置関係等に言及するときは、特に明示した場合および原理的に明らかにそうでないと考えられる場合等を除き、実質的にその形状等に近似または類似するもの等を含むものとする。

40

【 0 0 2 6 】

< 本発明の一実施形態に係るコンテンツ生成システムの概要 >

本発明の一実施形態に係るコンテンツ生成システムは、発話者による発話音声を入力として、それに対応する動画画像、漫画等のコンテンツを生成するものである。

【 0 0 2 7 】

発話者は、例えば、物語、ゲームプラン、体験、思い出等を話したり、小説等の書籍を朗読したりして音声を発話するものとする。発話者は、一人でもよいし、複数でもよい。

【 0 0 2 8 】

50

図 1 は、本発明の一実施形態に係るコンテンツ生成システム 10 の構成例を示している。

【0029】

コンテンツ生成システム 10 は、ネットワーク 11 を介して接続される端末装置 20 及びサーバ装置 30 を備える。

【0030】

ネットワーク 11 は、LAN (Local Area Network)、WAN (Wide Area Network)、インターネット、公衆電話通信網等からなる。

【0031】

端末装置 20 は、例えば、スマートフォン、タブレット型コンピュータ等の一般的なコンピュータから成る。

10

【0032】

端末装置 20 は、コンテンツの基となる発話音声を取得し、音声情報としてネットワーク 11 を介してサーバ装置 30 に送信する。また、端末装置 20 は、サーバ装置 30 からネットワーク 11 を介して送信されるコンテンツを受信、再生して表示する。

【0033】

サーバ装置 30 は、所謂クラウドネットワーク上に配置される。サーバ装置 30 は、サーバコンピュータ等の一般的なコンピュータから成る。サーバ装置 30 は、端末装置 20 から送信される発話音声に基づいてコンテンツを生成する。

【0034】

20

< 端末装置 20 を構成する機能ブロックの構成例 >

次に、図 2 は、端末装置 20 を構成する機能ブロックの構成例を示している。端末装置 20 は、制御部 21、通信部 22、音声取得部 23、音声出力部 24、操作部 25、及び表示部 26 の各機能ブロックを備える。

【0035】

制御部 21 は、例えば、コンピュータ等が内蔵する CPU (Central Processing Unit) が所定のプログラムを実行することによって実現される。制御部 21 は、端末装置 20 の全体の動作を制御する。

【0036】

通信部 22 は、例えば、コンピュータ等が内蔵する通信装置によって実現される。通信部 22 は、ネットワーク 11 を介してサーバ装置 30 に接続し、各種の情報を通信する。例えば、通信部 22 は、音声取得部 23 が取得した発話者の音声情報をサーバ装置 30 に送信する。また例えば、通信部 22 は、サーバ装置 30 から送信される、コンテンツ (動画像や漫画) を受信する。

30

【0037】

音声取得部 23 は、例えば、コンピュータ等が内蔵する出力装置に含まれるマイクロフォンによって実現される。音声取得部 23 は、発話者の音声を取得し、その結果得られる音声情報を通信部 22 に出力する。なお、音声取得部 23 によって取得した発話者の音声を録音し、発話後にサーバ装置 30 に送信してもよい。また、端末装置 20 に対して外部から発話者の音声を予め録音した音声情報を入力し、サーバ装置 30 に送信するようにしてもよい。

40

【0038】

音声出力部 24 は、例えば、コンピュータ等が内蔵する出力装置に含まれるスピーカによって実現される。音声出力部 24 は、例えば、サーバ装置 30 から送信されるコンテンツの音声を出力する。

【0039】

表示部 26 は、例えば、コンピュータ等が内蔵する出力装置に含まれるディスプレイによって実現される。表示部 26 は、サーバ装置 30 から供給されるコンテンツの画面を表示する。また、表示部 26 は、サーバ装置 30 から供給される画面情報に基づき、シナリオ情報 331 (図 4) や対応情報 333 (図 6) を修正するための修正画面を表示する。

50

【 0 0 4 0 】

<サーバ装置 3 0 を構成する機能ブロックの構成例>

次に、図 3 は、サーバ装置 3 0 を構成する機能ブロックの構成例を示している。サーバ装置 3 0 は、制御部 3 1、通信部 3 2、及び記憶部 3 3 の各機能ブロックを備える。

【 0 0 4 1 】

制御部 3 1 は、例えば、サーバコンピュータが内蔵する CPU が所定のプログラムを実行することにより実現される。制御部 3 1 は、サーバ装置 3 0 の全体を制御する。制御部 3 1 は、変換部 3 1 1、解析部 3 1 2、シナリオ情報生成部 3 1 3、対応付け部 3 1 4、動画像生成部 3 1 5、修正部 3 1 6、収集部 3 1 7、漫画生成部 3 1 8、及び表示制御部 3 1 9 の各機能ブロックを有する。なお、制御部 3 1 の各機能ブロックの少なくとも一部は、例えば機械学習や深層学習により得られた学習モデルを用いて実現してもよい。

10

【 0 0 4 2 】

変換部 3 1 1 は、端末装置 2 0 から送信される音声情報をテキスト情報に変換する。

【 0 0 4 3 】

解析部 3 1 2 は、音声情報及びテキスト情報を解析し、解析結果を記憶する。具体的には例えば、解析部 3 1 2 は、音声情報から発話者の発話音声のニュアンス（抑揚、強弱、トーン、スピード等）を検出したり、テキスト情報に自然言語処理を行ったりする。なお、テキスト情報の解析は、自然言語処理に限らず、任意の方法を適用できる。そして、解析部 3 1 2 は、解析結果としてコンテンツの 5 W 1 H を特定する。

【 0 0 4 4 】

具体的には例えば、発話音声のニュアンス、及びテキスト情報の自然言語処理の結果の少なくとも一方に基づき、コンテンツのタイトル、登場人物（語り手を含む）の感情（喜怒哀楽等）、発言（台詞）、状況（日時、季節、時代、場所等）、挙動、メタデータ（年代、性別、出身国（使用言語）、出身地方（使用言語の訛）等）を特定する。

20

【 0 0 4 5 】

さらに、解析部 3 1 2 は、発話音声のニュアンスに基づき、発話者のメタデータを特定する。なお、解析部 3 1 2 によって特定されたコンテンツのタイトル、登場人物、発話者のメタデータについては、ユーザが入力できるようにしてもよい。

【 0 0 4 6 】

シナリオ情報生成部 3 1 3 は、解析部 3 1 2 による解析結果に基づき、サーバ装置 3 0 にて生成されるコンテンツの脚本に相当するシナリオ情報 3 3 1 を生成し、記憶部 3 3 に格納する。

30

【 0 0 4 7 】

図 4 は、シナリオ情報 3 3 1 のデータ構造の一例を示している。同図の場合、シナリオ情報 3 3 1 には、タイトルが付与され、時系列 0 番として、発話者のメタデータが配置される。また、シナリオ情報 3 3 1 には、コンテンツの進行に対して時系列順（時系列 1 番以降）に登場人物の発言（台詞）及び挙動の少なくとも一方が配置される。さらに、シナリオ情報 3 3 1 には、登場人物が発言や挙動に行ったときの状況、及び感情が対応付けられている。さらに、シナリオ情報 3 3 1 には、登場人物のメタデータが配置される。なお、シナリオ情報 3 3 1 には、登場人物に関する少なくとも発言、挙動、状況、及び感情のうちの一つを時系列に配置すればよい。

40

【 0 0 4 8 】

図 3 に戻る。対応付け部 3 1 4 は、生成されたシナリオ情報 3 3 1 を、その内容に応じて、登場人物（語り手を含む）の発言及び挙動を少なくとも 1 以上含むように所定の編集単位毎（例えば、シーン毎等）に区分する。さらに、対応付け部 3 1 4 は、素材画像 DB（データベース）3 3 2 に格納されている素材画像のうち、シナリオ情報 3 3 1 の編集単位毎に、登場人物の感情、状況、挙動、メタデータに一致するもの（完全に一致するものに限らず、類似しているものを含む）を検索する。そして、対応付け部 3 1 4 は、各編集単位に素材画像を対応付けた対応情報 3 3 3 を生成し、記憶部 3 3 に格納する。

【 0 0 4 9 】

50

なお、素材画像の検索に際しては、発話者のメタデータに、素材画像における登場人物のメタデータが一致するものを検索するようにしてもよい。具体的には例えば、発話者の出身国が日本である場合、登場人物には日本人が登場し、日本語を話している素材画像を検出するようにする。また、発話者のメタデータに、素材画像の作風が一致するものを検索するようにしてもよい。具体的には例えば、発話者の女性である場合、作風が女性向けである素材画像を検出するようにする。なお、ユーザが、検索する素材画像の作風を選択できるようにしてもよい。そして、複数の素材画像が検索された場合、その中からシナリオ情報 3 3 1 の各編集単位に対応付けるものを対応付け部 3 1 4 が選択してもよいし、ユーザに選択させてもよい。

【 0 0 5 0 】

図 5 は、素材画像 D B 3 3 2 のデータ構造の一例を示している。素材画像 D B 3 3 2 は予め生成されており、記憶部 3 3 に格納されている。素材画像 D B 3 3 2 には、素材 I D に対応付けて、素材画像データ、素材内容、パラメータ、メタデータ、及び作風が記録されている。素材 I D は、素材画像を特定するための識別子である。素材画像データは、例えば、3 D モーションデータや 2 D モーションデータの動画像であり、そのデータ形式は任意である。また、素材画像データは、静止画像であり、そのデータ形式は任意である。素材内容には、素材画像データの内容（登場人物とその挙動等）が記録されている。パラメータには、素材画像データにおける登場人物の感情（喜怒哀楽等）、状況（時刻、季節、時代、場所）等が数値化されて記録されている。メタデータには、素材画像データにおける登場人物の年代、性別、出身国（使用言語）、出身地方（使用言語の訛）等のメタデータが記録されている。作風には、素材画像データの作風（万人向け、男性向け、女性向け、子供向け、成人向け、時代劇、西洋風、東洋風、実写、アニメーション等）が記録されている。

【 0 0 5 1 】

図 6 は、対応情報 3 3 3 のデータ構造の一例を示している。対応情報 3 3 3 には、シナリオ情報 3 3 1 の各編集単位の時系列番号（図 6 の場合、シーン番号）に対応付けて素材 I D が記録される。

【 0 0 5 2 】

図 3 に戻る。動画像生成部 3 1 5 は、生成された対応情報 3 3 3 に基づき、素材 I D に対応する素材画像データを素材画像 D B 3 3 2 から取得して時系列順に連結する。さらに、動画像生成部 3 1 5 は、シナリオ情報 3 3 1 を参照し、登場人物の発話（台詞）を、連結した素材画像データに字幕として重畳したり、あるいは発話に対応する音声を合成したりして、字幕または合成音声を含むコンテンツとしての動画像を生成して記憶部 3 3 に格納する。

【 0 0 5 3 】

修正部 3 1 6 は、端末装置 2 0 から送信されるユーザからの修正指示に基づき、シナリオ情報 3 3 1 を修正、補完する。また、修正部 3 1 6 は、端末装置 2 0 から送信されるユーザからの修正指示に基づき、対応情報 3 3 3 を修正する。なお、ユーザが対応情報 3 3 3 を修正する場合、ユーザが任意の素材画像データ（動画像または静止画像のどちらでもよい）を素材画像 D B 3 3 2 に追加登録して、シナリオ情報 3 3 1 の任意の編集単位に対応付けるようにしてもよい。さらに、修正部 3 1 6 は、シナリオ情報 3 3 1 や対応情報 3 3 3 に対する修正結果を学習し、学習結果に基づいて素材画像 D B 3 3 2 における素材内容やパラメータを更新する。

【 0 0 5 4 】

収集部 3 1 7 は、インターネットにて、所謂、フリー素材として公開されている画像データ（動画像データ、及び静止画像データ）を収集し、収集した画像データを解析して、その素材内容、パラメータ、メタデータ、及び作風を設定し、素材画像データとして素材画像 D B 3 3 2 に登録する。

【 0 0 5 5 】

漫画生成部 3 1 8 は、動画像生成部 3 1 5 によって生成された動画像に基づき、コンテ

10

20

30

40

50

ンツとしての漫画を生成して記憶部 3 3 に格納する。具体的には例えば、動画像の各シーンから代表画像を抽出し、代表画像に登場人物の発話（台詞）を吹き出しとして重畳することにより、コンテンツとしての漫画を生成する。

【 0 0 5 6 】

表示制御部 3 1 9 は、端末装置 2 0 からの要求に応じ、生成されたコンテンツを端末装置 2 0 に供給して再生、表示させる。また、表示制御部 3 1 9 は、端末装置 2 0 からの要求に応じ、シナリオ情報 3 3 1 や対応情報 3 3 3 をユーザが修正可能な修正画面を表示するための画面情報を生成する。生成された画面情報は、通信部 3 2 によって端末装置 2 0 に送信される。

【 0 0 5 7 】

通信部 3 2 は、例えば、サーバコンピュータが内蔵する通信装置により実現される。通信部 3 2 は、ネットワーク 1 1 を介して接続してきた端末装置 2 0 と各種の情報を通信する。例えば、通信部 3 2 は、端末装置 2 0 から送信される音声情報を受信する。また、例えば、通信部 3 2 は、ネットワーク 1 1 を介し、端末装置 2 0 にコンテンツ（動画像や漫画）や画面情報を送信する。

【 0 0 5 8 】

記憶部 3 3 は、例えばサーバコンピュータが内蔵する記憶装置によって実現される。記憶部 3 3 には、シナリオ情報 3 3 1、素材画像 DB 3 3 2、対応情報 3 3 3、及びコンテンツ 3 3 4 が格納される。記憶部 3 3 に格納されたコンテンツ 3 3 4 は、コンテンツ 3 3 4 の基となった発話音声を送信してきた端末装置 2 0 の他、所定の端末装置等に供給するようにしてもよい。

【 0 0 5 9 】

なお、図 3 に示されたサーバ装置 3 0 の機能ブロックの一部または全てを端末装置 2 0 に設けるようにしてもよい。

【 0 0 6 0 】

< コンテンツ生成システム 1 0 によるコンテンツ生成処理 >

次に、コンテンツ生成システム 1 0 によるコンテンツ生成処理について説明する。図 7 は、コンテンツ生成システム 1 0 によるコンテンツ生成処理の一例を説明するフローチャートである。

【 0 0 6 1 】

該コンテンツ作成処理は、例えば、端末装置 2 0 に対するユーザ（発話者等）からの所定の開始操作に応じて開始される。

【 0 0 6 2 】

はじめに、端末装置 2 0 は、音声取得部 2 3 により、発話者の発話音声を取得し、その結果得られる音声情報を通信部 2 2 に出力し、通信部 2 2 により、音声取得部 2 3 からの音声情報を、ネットワーク 1 1 を介してサーバ装置 3 0 に送信する。サーバ装置 3 0 は、通信部 3 2 により、端末装置 2 0 からの音声情報を制御部 3 1 に出力する（ステップ S 1）。

【 0 0 6 3 】

次に、制御部 3 1 は、変換部 3 1 1 により、発話者の音声情報をテキスト情報に変換する（ステップ S 2）。

【 0 0 6 4 】

次に、制御部 3 1 は、解析部 3 1 2 により、音声情報及びテキスト情報を解析する（ステップ S 3）。次に、制御部 3 1 は、シナリオ情報生成部 3 1 3 により、解析部 3 1 2 による解析結果に基づいてシナリオ情報 3 3 1 を生成し、記憶部 3 3 に格納する（ステップ S 4）。

【 0 0 6 5 】

次に、制御部 3 1 は、対応付け部 3 1 4 により、シナリオ情報 3 3 1 を所定の編集単位毎に区分し（ステップ S 5）、素材画像 DB 3 3 2 を参照し、シナリオ情報 3 3 1 の編集単位毎に素材画像を対応付けた対応情報 3 3 3 を生成し、記憶部 3 3 に格納する（ステッ

10

20

30

40

50

プS6)。

【0066】

次に、制御部31は、動画像生成部315により、対応情報333に基づき、素材IDに対応する素材画像データを素材画像DB332から取得して時系列順に連結する。さらに、動画像生成部315は、シナリオ情報331を参照し、登場人物の発話(台詞)を、連結した素材画像データに字幕として重畳したり、あるいは発話に対応する音声を合成したりして、字幕または合成音声を含む動画像を生成し、コンテンツ334として記憶部33に格納する(ステップS7)。

【0067】

次に、制御部31は、漫画生成部318により、動画像に基づいて漫画を生成し、コンテンツ334として記憶部33に格納する(ステップS8)。

【0068】

以上で、コンテンツ生成システム10によるコンテンツ生成処理は終了される。なお、生成されたコンテンツ334は、端末装置20等からの要求に応じて要求元に供給されて再生される。

【0069】

以上に説明したコンテンツ生成処理によれば、発話者による発話音声に基づいてコンテンツとして動画像及び漫画を生成することが可能となる。

【0070】

<コンテンツ生成システム10による修正処理>

次に、コンテンツ生成システム10による修正処理について説明する。図8は、コンテンツ生成システム10による修正処理の一例を説明するフローチャートである。

【0071】

該修正処理は、端末装置20に対してユーザから修正要求が入力され、該修正要求がサーバ装置30に送信されて制御部31に入力されたときに開始される。

【0072】

はじめに、制御部31は、修正部316により、ユーザからの修正要求がシナリオ情報331の修正を要求するものであるか否かを判定する(ステップS11)。

【0073】

ここで、修正部316がシナリオ情報331の修正を要求するものであると判定した場合(ステップS11でYES)、処理はステップS12に進められる。次に、制御部31は、表示制御部319により、シナリオ情報331を修正するための修正画面の画面情報を生成し、通信部32が、生成された画面情報を端末装置20に送信する。端末装置20では、送信された画面情報に基づいてシナリオ情報331を修正するための画面が表示され、該画面に対し、操作部25を用いてユーザが修正指示を入力すると、通信部22が該修正指示をサーバ装置30に送信する。そして、サーバ装置30にて、制御部31は、修正部316により、ユーザからの修正指示に応じて、記憶部33に格納されているシナリオ情報331を修正、補完する(ステップS12)。

【0074】

なお、修正部316がシナリオ情報331の修正を要求するものではないと判定した場合(ステップS11でNO)、ステップS12はスキップされ、処理はステップS13に進められる。

【0075】

次に、制御部31は、修正部316により、ユーザからの修正要求が対応情報333の修正を要求するものであるか否かを判定する(ステップS13)。

【0076】

ここで、修正部316が対応情報333の修正を要求するものであると判定した場合(ステップS13でYES)、処理はステップS14に進められる。次に、制御部31は、表示制御部319により、対応情報333を修正するための修正画面の画面情報を生成し、通信部32が、生成された画面情報を端末装置20に送信する。端末装置20では、送

10

20

30

40

50

信された画面情報に基づいて対応情報 3 3 3 を修正するための画面が表示され、該画面に対し、操作部 2 5 を用いてユーザが修正指示を入力すると、通信部 2 2 が該修正指示をサーバ装置 3 0 に送信する。そして、サーバ装置 3 0 にて、制御部 3 1 は、修正部 3 1 6 により、ユーザからの修正指示に応じて、記憶部 3 3 に格納されている対応情報 3 3 3 を修正する（ステップ S 1 4 ）。

【 0 0 7 7 】

なお、修正部 3 1 6 が対応情報 3 3 3 の修正を要求するものではないと判定した場合（ステップ S 1 3 で N O ）、ステップ S 1 4 はスキップされる。

【 0 0 7 8 】

次に、制御部 3 1 は、動画像生成部 3 1 5 により、修正されたシナリオ情報 3 3 1 及び対応情報 3 3 3 に基づいてコンテンツとしての動画像を再生成して記憶部 3 3 に格納するとともに、漫画生成部 3 1 8 により、再生成された動画像に基づいて、コンテンツとしての漫画を再生成して記憶部 3 3 に格納する（ステップ S 1 5 ）。

【 0 0 7 9 】

次に、制御部 3 1 は、修正部 3 1 6 により、シナリオ情報 3 3 1 や対応情報 3 3 3 に対する修正結果を学習し、学習結果に基づいて素材画像 D B 3 3 2 における素材内容やパラメータを更新する（ステップ S 1 6 ）。なお、シナリオ情報生成部 3 1 3 や対応付け部 3 1 4 においても、シナリオ情報 3 3 1 や対応情報 3 3 3 に対する修正結果を学習し、以降の処理に適用するようにしてもよい。以上で、コンテンツ生成システム 1 0 による修正処理は終了される。

【 0 0 8 0 】

コンテンツ生成システム 1 0 による修正処理によれば、ユーザはシナリオ情報 3 3 1 や対応情報 3 3 3 を修正し、修正結果を反映させたコンテンツを得ることができる。さらに、ユーザからの修正指示に基づく修正結果が学習されて素材画像 D B 3 3 2 に反映されるので、これ以降、より適切な素材画像が検索されてコンテンツが生成されるようになる。

【 0 0 8 1 】

< コンテンツ表示画面について >

次に、図 9 は、端末装置 2 0 等におけるコンテンツ表示画面 5 0 の表示例を示している。

【 0 0 8 2 】

コンテンツ表示画面 5 0 には、コンテンツが表示されるコンテンツ再生領域 5 1 が設けられている。コンテンツ再生領域 5 1 には、字幕 5 3 を表示させることができる。コンテンツ再生領域 5 1 の上側にはコンテンツのタイトル 5 2 が表示される。コンテンツ再生領域 5 1 の下側には、ユーザがコンテンツの再生開始を指示するための操作ボタン 5 4、コンテンツの早戻しを指示するための操作ボタン 5 5、及びコンテンツの早送りを指示するための操作ボタン 5 6 が設けられている。なお、タイトル 5 2、字幕 5 3、及び操作ボタン 5 4 ~ 5 6 の表示位置は、図示した例に限られず任意である。

【 0 0 8 3 】

コンテンツ再生領域 5 1 の右側には、コンテンツの現在の作風を表すとともに、ユーザがコンテンツの作風の変更を指示することができる作風ボタン 6 1 ~ 6 5 が設けられている。同図の場合、作風ボタン 6 1 ~ 6 5 のうち、作風（女性向け）ボタン 6 2 が強調表示されており、コンテンツの現在の作風が女性向けであることを表している。

【 0 0 8 4 】

この状態において、例えば、ユーザが作風（子供向け）ボタン 6 3 を選択した場合、上述したコンテンツ生成処理（図 7 ）のステップ S 6 ~ S 8 が再度実行されて、子供向けの素材画像データを用いてコンテンツが再生成され、コンテンツ再生領域 5 1 に表示される。

【 0 0 8 5 】

なお、作風ボタン 6 1 ~ 6 5 の表示位置や数は、図示した例に限られず任意である。また、作風ボタンを 1 つだけ表示し、該作風ボタンが操作された場合、作風の選択肢がポッ

10

20

30

40

50

プアップ表示されるようにしてもよい。

【 0 0 8 6 】

< 一般的なコンピュータの構成例 >

上述したように、端末装置 2 0、及びサーバ装置 3 0 は、一般的なコンピュータによって実現することができる。

【 0 0 8 7 】

図 1 0 は、端末装置 2 0、及びサーバ装置 3 0 を実現する一般的なコンピュータの構成例を示している。

【 0 0 8 8 】

該コンピュータ 1 0 0 において、CPU 1 0 1、ROM (Read Only Memory) 1 0 2、RAM (Random Access Memory) 1 0 3 は、バス 1 0 4 により相互に接続されている。

10

【 0 0 8 9 】

バス 1 0 4 には、さらに、入出力インターフェース 1 0 5 が接続されている。入出力インターフェース 1 0 5 には、入力装置 1 0 6、出力装置 1 0 7、記憶装置 1 0 8、通信装置 1 0 9、およびドライブ装置 1 1 0 が接続されている。

【 0 0 9 0 】

入力装置 1 0 6 は、キーボード、マウス、タッチパネル、マイクロフォン等よりなり、例えば、端末装置 2 0 の音声取得部 2 3 及び操作部 2 5 (図 2) として機能する。出力装置 1 0 7 は、ディスプレイ、スピーカ等よりなり、例えば、端末装置 2 0 の音声出力部 2 4 及び表示部 2 6 (図 2) として機能する。

20

【 0 0 9 1 】

記憶装置 1 0 8 は、HDD (hard Disk Drive)、SSD (solid State Drive) などからなり、例えば、サーバ装置 3 0 の記憶部 3 3 (図 3) として機能する。通信装置 1 0 9 は、LAN インターフェースや Wi-Fi インターフェースからなり、例えば、端末装置 2 0 の通信部 2 2 (図 2)、及びサーバ装置 3 0 の通信部 3 2 (図 3) として機能する。ドライブ装置 1 1 0 は、磁気ディスク、光ディスク、光磁気ディスク、又は半導体メモリ等のリムーバブルメディア 1 1 1 を駆動する。

【 0 0 9 2 】

以上のように構成されるコンピュータ 1 0 0 では、CPU 1 0 1 が、例えば、記憶装置 1 0 8 に記憶されているプログラムを、入出力インターフェース 1 0 5 およびバス 1 0 4 を介して、RAM 1 0 3 にロードして実行することにより、例えば、端末装置 2 0 の制御部 2 1 (図 2)、及びサーバ装置 3 0 の制御部 3 1 (図 3) が実現される。

30

【 0 0 9 3 】

コンピュータ 1 0 0 (CPU 1 0 1) が実行するプログラムは、例えば、パッケージメディア等としてのリムーバブルメディア 1 1 1 に記録して提供することができる。また、プログラムは、ローカルエリアネットワーク、インターネット、デジタル衛星放送といった、有線または無線の伝送媒体を介して提供することができる。

【 0 0 9 4 】

コンピュータ 1 0 0 では、プログラムは、リムーバブルメディア 1 1 1 をドライブ装置 1 1 0 に装着することにより、入出力インターフェース 1 0 5 を介して、記憶装置 1 0 8 にインストールすることができる。また、プログラムは、有線または無線の伝送媒体を介して、通信装置 1 0 9 で受信し、記憶装置 1 0 8 にインストールすることができる。その他、プログラムは、ROM 1 0 2 や記憶装置 1 0 8 に、あらかじめインストールしておくことができる。

40

【 0 0 9 5 】

なお、コンピュータ 1 0 0 が実行するプログラムは、本明細書で説明する順序に沿って時系列に処理が行われるプログラムであってもよいし、並列に、あるいは呼び出しが行われたとき等の必要なタイミングで処理が行われるプログラムであってもよい。

【 0 0 9 6 】

本明細書に記載された効果はあくまで例示であって限定されるものではなく、他の効果

50

があってもよい。

【0097】

本発明は、上記した実施形態に限定されるものではなく、様々な変形例が含まれる。例えば、上記した各実施形態は、本発明を分かりやすく説明するために詳細に説明したものであり、本発明が、必ずしも説明した全ての構成要素を備えるものに限定されるものではない。また、ある実施形態の構成の一部を、他の実施形態の構成に置き換えることが可能であり、ある実施形態の構成に、他の実施形態の構成を加えることも可能である。また、各実施形態の構成の一部について、他の構成の追加・削除・置換をすることが可能である。

【0098】

また、上記の各構成、機能、処理部、処理手段等は、それらの一部または全部を、例えば集積回路で設計する等によりハードウェアで実現してもよい。また、上記の各構成、機能等は、プロセッサがそれぞれの機能を実現するプログラムを解釈し、実行することによりソフトウェアで実現されてもよい。各機能を実現するプログラム、テーブル、ファイル等の情報は、半導体メモリ、HDD、SSD等の記憶装置、または、ICカード、SDカード、DVD等の記録媒体に置くことができる。また、制御線や情報線は説明上必要と考えられるものを示しており、製品上必ずしも全ての制御線や情報線を示しているとは限らない。実際には殆ど全ての構成が相互に接続されていると考えてもよい。

【符号の説明】

【0099】

10・・・コンテンツ生成システム、11・・・ネットワーク、20・・・端末装置、21・・・制御部、22・・・通信部、23・・・音声取得部、24・・・音声出力部、25・・・操作部、26・・・表示部、30・・・サーバ装置、31・・・制御部、32・・・通信部、33・・・記憶部、50・・・コンテンツ表示画面、51・・・コンテンツ再生領域、52・・・タイトル、53・・・字幕、54～56・・・操作ボタン、61～65・・・作風ボタン、100・・・コンピュータ、101・・・CPU、102・・・ROM、103・・・RAM、104・・・バス、105・・・入出力インターフェース、106・・・入力装置、107・・・出力装置、108・・・記憶装置、109・・・通信装置、110・・・ドライブ装置、111・・・リムーバブルメディア、311・・・変換部、312・・・解析部、313・・・シナリオ情報生成部、314・・・対応付け部、315・・・動画像生成部、316・・・修正部、317・・・収集部、318・・・漫画生成部、319・・・表示制御部、331・・・シナリオ情報、332・・・素材画像DB、333・・・対応情報、334・・・コンテンツ

【要約】

【課題】発話音声に基づいてコンテンツを生成する。

【解決手段】コンテンツ生成システムは、発話者の音声情報を取得する音声取得部と、前記音声情報をテキスト情報に変換する変換部と、前記テキスト情報を解析する解析部と、前記テキスト情報の解析結果に基づいてシナリオ情報を生成するシナリオ情報生成部と、素材画像が格納された素材画像データベースと、前記シナリオ情報を所定の編集単位毎に区分し、前記編集単位に前記素材画像データベースに格納された前記素材画像を対応付けて対応情報を生成する対応付け部と、前記編集単位毎に対応付けられた前記素材画像を連結して動画像を生成する動画像生成部と、を備えることを特徴とする。

【選択図】図3

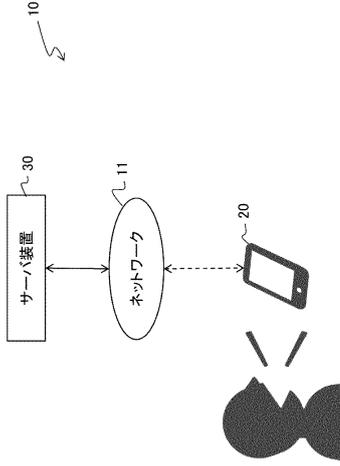
10

20

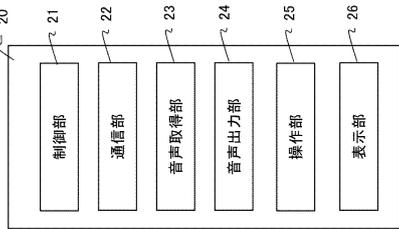
30

40

【 図 1 】



【 図 2 】



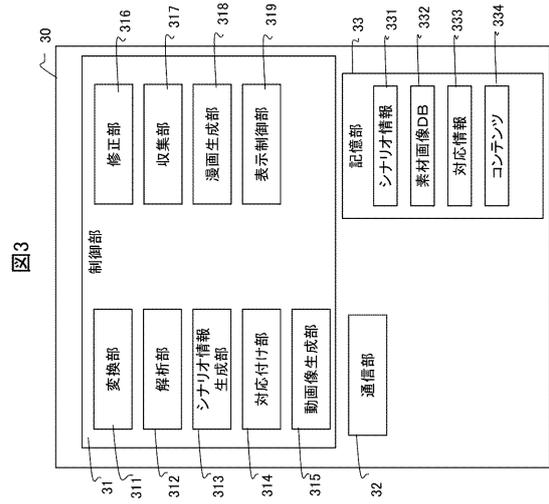
【 図 4 】

図4

タイトル	桃太郎	登場人物	発言	奉勳	状況	感情	メタデータ
0	発話者	-	-	-	-	-	男性 50代 日本 東北地方
1	語り手	昔々、お爺さんとお婆さんが...	-	-	女性 20代
2	お爺さん	-	薪拾い	山	疲れ	男性 老人	女性 老人
3	お婆さん	大きな桃だ	洗濯中に桃を拾う	川	驚き	女性 老人	女性 老人
4	お婆さん	切ってみましょう	桃を切る	家	喜び	女性 老人	女性 老人
5	語り手	桃太郎と名付けて大切に育てました	-	-	喜び	女性 20代	女性 20代
⋮							

331

【 図 3 】



【 図 5 】

図5

素材ID	素材画像データ	素材内容	ハラメータ	メタデータ	作風
abc11wmv	人が歩いている	女性向き
abc12mp4	複数の人が会議をしている	男性向き
abc13gif	猫がじゃれている	子供向き
abc14wmv	兵士が剣で戦っている	男性向き
abc15mpg	人が電話をしている	女性向き
⋮					

332

【 図 6 】

シーン番号	素材ID
シーン1	abc12
シーン2	xvc61
シーン3	abc72
シーン4	xyz22
シーン5	bbb773
⋮	

【 図 7 】

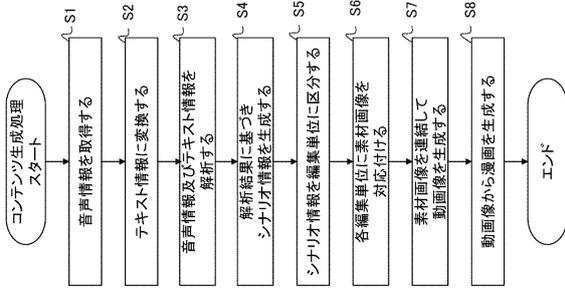


図7

【 図 9 】

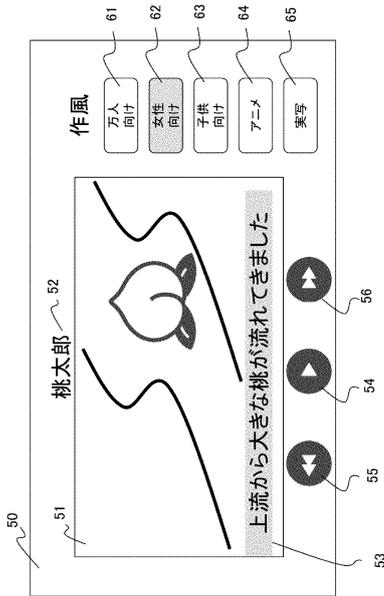


図9

【 図 8 】

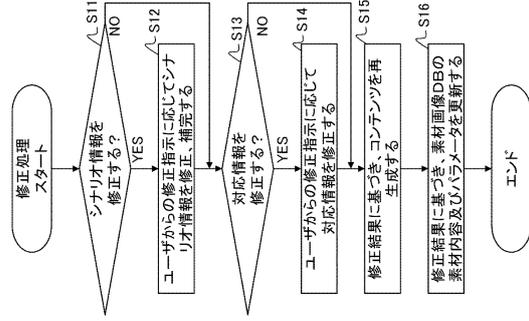


図8

【 図 10 】

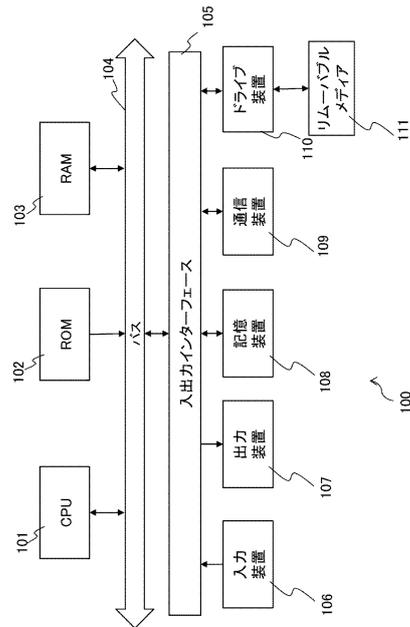


図10

フロントページの続き

- (56)参考文献 米国特許出願公開第2006/0217979(US, A1)
国際公開第2007/138911(WO, A1)
特開2002-091482(JP, A)
特開2001-147752(JP, A)
国際公開第2006/059570(WO, A1)
特開2011-002875(JP, A)

(58)調査した分野(Int.Cl., DB名)

G06T 13/00 - 13/80