(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7621607号
(P7621607)

(45)発行日　令和7年1月27日(2025.1.27)

(24)登録日　令和7年1月17日(2025.1.17)

(51)国際特許分類

		FI	
H04N	*21/435(2011.01)*	H04N	21/435
H04N	*21/854(2011.01)*	H04N	21/854

請求項の数　17　（全35頁）

(21)出願番号　特願2023-547248(P2023-547248)
(86)(22)出願日　令和4年2月8日(2022.2.8)
(65)公表番号　特表2024-505988(P2024-505988A)
(43)公表日　令和6年2月8日(2024.2.8)
(86)国際出願番号　PCT/US2022/070572
(87)国際公開番号　WO2022/170368
(87)国際公開日　令和4年8月11日(2022.8.11)
審査請求日　令和5年8月3日(2023.8.3)
(31)優先権主張番号　17/170,695
(32)優先日　令和3年2月8日(2021.2.8)
(33)優先権主張国・地域又は機関
米国(US)

(73)特許権者　000002185
ソニーグループ株式会社
東京都港区港南１丁目７番１号
(74)代理人　100092093
弁理士　辻居 幸一
(74)代理人　100109070
弁理士　須田 洋之
(74)代理人　100067013
弁理士　大塚 文昭
(74)代理人
上杉 浩
(74)代理人　100120525
弁理士　近藤 直樹
(72)発明者　キャンデロア ブラント
アメリカ合衆国 カリフォルニア州 ９２
最終頁に続く

(54)【発明の名称】　シーン説明の再生制御

(57)【特許請求の範囲】

【請求項１】

メディアレンダリング装置であって、

一連の撮影シーンと、前記一連の撮影シーンにおける撮影シーンを説明するテキストベースのビデオ説明情報及びタイミング情報を含むテキスト情報とを含むメディアコンテンツを取得し、

前記撮影シーンの前記テキスト情報から、前記ビデオ説明情報を再生するためのタイミング情報を抽出し、

前記撮影シーンの連続する音声部分間の自然な休止に対応する第１の時間間隔を前記タイミング情報から抽出し、

前記一連の撮影シーンにおける前記撮影シーンのオーディオ部分を再生するための時間間隔をそれぞれが示す、前記撮影シーンの一連の第２の時間間隔を決定し、

前記撮影シーンの前記ビデオ説明情報のオーディオ表現の第３の時間間隔を決定し、

前記決定された第３の時間間隔と前記第１の時間間隔の比率に基づいて乗数を決定し、

前記乗数と前記ビデオ説明情報の実際の再生速度とに基づいて、前記ビデオ説明情報のオーディオ表現を再生する速度を決定し、

前記決定された速度に基づいて、前記ビデオ説明情報の再生を、前記撮影シーンの前記抽出されたタイミング情報によって示される第１の時間間隔においてテキスト表現又はテキスト及びオーディオ表現で制御する、

ように構成された回路を備える、

ことを特徴とするメディアレンダリング装置。
【請求項２】
前記回路は、
それぞれが前記一連の第２の時間間隔とは異なる、前記撮影シーンの一連の第４の時間間隔を決定し、
前記一連の第４の時間間隔から、時間間隔閾値よりも高い前記第１の時間間隔を選択する、
ようにさらに構成される、請求項１に記載のメディアレンダリング装置。
【請求項３】
前記決定される速度は、前記オーディオ表現の実際の再生速度よりも低い、
請求項１に記載のメディアレンダリング装置。
【請求項４】
前記決定される速度は、前記オーディオ表現の実際の再生速度よりも高い、
請求項１に記載のメディアレンダリング装置。
【請求項５】
前記回路は、前記メディアレンダリング装置に関連する規定の速度設定に基づいて、前記ビデオ説明情報の前記オーディオ表現を再生する前記速度を決定するようにさらに構成され、
前記規定の速度設定は、前記ビデオ説明情報の前記オーディオ表現の最大再生速度を示す、
請求項１に記載のメディアレンダリング装置。
【請求項６】
前記回路は、前記テキスト情報と共に速度情報を受け取り、前記決定された速度及び前記規定の速度設定に基づいて、前記撮影シーンの画像部分又はオーディオ部分の一方の再生を制御するようにさらに構成される、
請求項５に記載のメディアレンダリング装置。
【請求項７】
前記回路は、
前記メディアコンテンツがレンダリングされている対象であるユーザのプロファイル情報を示す第２のユーザ入力を受け取り、
前記受け取った第２のユーザ入力に基づいて、前記ビデオ説明情報の前記オーディオ表現を再生するための速度設定を決定する、
ようにさらに構成される、請求項５に記載のメディアレンダリング装置。
【請求項８】
前記回路は、
前記一連の撮影シーンのうちの１つの撮影シーンの説明に対応する第１のユーザ入力を受け取り、
前記受け取った第１のユーザ入力を、前記一連の撮影シーンの各々に関連する前記ビデオ説明情報内で検索し、
前記検索に基づいて、前記メディアコンテンツを再生するための再生タイミング情報を決定し、
前記決定された再生タイミング情報に基づいて前記メディアコンテンツの前記再生を制御する、
ようにさらに構成される、請求項１に記載のメディアレンダリング装置。
【請求項９】
前記第１の時間間隔は、前記撮影シーンの第１のせりふと第２のせりふとの間である、
請求項１に記載のメディアレンダリング装置。
【請求項１０】
前記第１のせりふは、前記撮影シーンの第１のショットの最後の単語であり、前記第２のせりふは、前記撮影シーンの第２のショットの最初の単語であり、

前記第１のショット及び前記第２のショットは、前記撮影シーンの連続するショットである、
請求項９に記載のメディアレンダリング装置。
【請求項１１】
訓練済み機械学習（ＭＬ）モデルを記憶するように構成されたメモリをさらに備え、前記テキスト情報は速度情報をさらに含み、前記回路は、
前記撮影シーンの少なくとも１つの特性の分析に基づいて前記撮影シーンのコンテキスト情報を決定し、
前記撮影シーンの前記決定されたコンテキスト情報に対する前記訓練済みＭＬモデルの適用に基づいて、前記ビデオ説明情報のオーディオ表現を再生するためのオーディオ特性を決定し、
前記速度情報及び前記決定されたオーディオ特性に基づいて、前記撮影シーンの前記抽出されたタイミング情報によって示される前記第１の時間間隔において前記ビデオ説明情報の前記オーディオ表現を再生するように制御する、
ようにさらに構成される、請求項１に記載のメディアレンダリング装置。
【請求項１２】
前記撮影シーンを説明する前記ビデオ説明情報は、前記撮影シーン内に存在する生物オブジェクト又は無生物オブジェクトに関する認知情報を含み、
前記回路は、前記撮影シーンの前記ビデオ説明情報に含まれる前記認知情報の再生を制御するようにさらに構成される、
請求項１に記載のメディアレンダリング装置。
【請求項１３】
前記メディアレンダリング装置は、前記ビデオ説明情報のテキスト表現を再生するように構成されたディスプレイ装置をさらに備える、
請求項１に記載のメディアレンダリング装置。
【請求項１４】
前記メディアコンテンツは、前記一連の撮影シーンの各々のオーディオ部分を表すクローズドキャプション情報をさらに含み、
前記一連の撮影シーンの各々を説明する前記ビデオ説明情報は、前記クローズドキャプション情報と共に前記メディアコンテンツ内に符号化される、
請求項１に記載のメディアレンダリング装置。
【請求項１５】
前記回路は、前記メディアレンダリング装置に関連するオーディオレンダリング装置を、前記ビデオ説明情報のオーディオ表現及び前記撮影シーンのオーディオ部分を再生するように制御するようさらに構成される、
請求項１に記載のメディアレンダリング装置。
【請求項１６】
メディアレンダリング装置において、
一連の撮影シーンと、前記一連の撮影シーンにおける撮影シーンを説明するテキストベースのビデオ説明情報及びタイミング情報を含むテキスト情報とを含むメディアコンテンツを取得することと、
前記撮影シーンの前記テキスト情報から、前記ビデオ説明情報を再生するためのタイミング情報を抽出することと、
前記撮影シーンの連続する音声部分間の自然な休止に対応する第１の時間間隔を前記タイミング情報から抽出することと、
前記一連の撮影シーンにおける前記撮影シーンのオーディオ部分を再生するための時間間隔をそれぞれが示す、前記撮影シーンの一連の第２の時間間隔を決定することと、
前記撮影シーンの前記ビデオ説明情報のオーディオ表現の第３の時間間隔を決定することと、
前記決定された第３の時間間隔と前記第１の時間間隔の比率に基づいて乗数を決定する

ことと、
前記乗数と前記ビデオ説明情報の実際の再生速度とに基づいて、前記ビデオ説明情報のオーディオ表現を再生する速度を決定することと、
前記決定された速度に基づいて、前記ビデオ説明情報の再生を、前記撮影シーンの前記抽出されたタイミング情報によって示される第１の時間間隔においてテキスト表現又はテキスト及びオーディオ表現で制御することと、
を含むことを特徴とする方法。
【請求項１７】
コンピュータ実行可能命令を記憶した非一時的コンピュータ可読媒体であって、前記コンピュータ実行可能命令は、メディアレンダリング装置によって実行された時に、前記メディアレンダリング装置に、
一連の撮影シーンと、前記一連の撮影シーンにおける撮影シーンを説明するテキストベースのビデオ説明情報及びタイミング情報を含むテキスト情報とを含むメディアコンテンツを取得することと、
前記撮影シーンの前記テキスト情報から、前記ビデオ説明情報を再生するためのタイミング情報を抽出することと、
前記撮影シーンの連続する音声部分間の自然な休止に対応する第１の時間間隔を前記タイミング情報から抽出することと、
前記一連の撮影シーンにおける前記撮影シーンのオーディオ部分を再生するための時間間隔をそれぞれが示す、前記撮影シーンの一連の第２の時間間隔を決定することと、
前記撮影シーンの前記ビデオ説明情報のオーディオ表現の第３の時間間隔を決定することと、
前記決定された第３の時間間隔と前記第１の時間間隔の比率に基づいて乗数を決定することと、
前記乗数と前記ビデオ説明情報の実際の再生速度とに基づいて、前記ビデオ説明情報のオーディオ表現を再生する速度を決定することと、
前記決定された速度に基づいて、前記ビデオ説明情報の再生を、前記撮影シーンの前記抽出されたタイミング情報によって示される第１の時間間隔においてテキスト表現又はテキスト及びオーディオ表現で制御することと、
を含む動作を実行させる、ことを特徴とする非一時的コンピュータ可読媒体。
【発明の詳細な説明】
【技術分野】
【０００１】
〔関連出願との相互参照／引用による組み入れ〕
なし
【０００２】
本開示の様々な実施形態は、メディア再生制御に関する。具体的には、本開示の様々な実施形態は、メディアレンダリング装置及びシーン説明の再生制御方法に関する。
【背景技術】
【０００３】
近年のメディアコンテンツ再生分野の進歩は、メディアコンテンツの様々な部分を制御するための様々な技術の発展をもたらした。通常、メディアコンテンツ（例えば、映画）は、１又は２以上の視聴者のためにメディアレンダリング装置上で同時にレンダリングできるビデオトラック及び対応するオーディオトラックなどの異なる部分を含むことができる。いくつかの状況では、視覚障害者又は認知障害者などの視聴者が、映画を理解する上でメディアコンテンツのシーンにおける要素、文脈、筋書き又は感情を視覚化できないという問題に直面することがある。メディアコンテンツの中には、視覚障害又は認知障害視聴者のメディアコンテンツ体験をさらに強化するために、ビデオトラック及びオーディオトラックと共に、メディアコンテンツ内にビデオ説明付きオーディオを代替オーディオトラックとして含むことができるものもある。いくつかのシナリオでは、ビデオ説明がオー

ディオベースであってビデオの説明に使用され、従って「ビデオ説明（ｖｉｄｅｏ　ｄｅｓｃｒｉｐｔｉｏｎ）」と呼ばれる。しかしながら、米国では、連邦通信委員会（ＦＣＣ）が、２０２０年１１月３０日に発行された２１世紀における通信とビデオアクセシビリティに関する２０１０年法、ＦＣＣ２０－１５５（２０２０年）によってこの用語を「オーディオ説明（ａｕｄｉｏ　ｄｅｓｃｒｉｐｔｉｏｎ）」に変更した。本文書では、古い用語である「ビデオ説明」を引き続き使用する。このナレーション付きの説明は、視覚障害者又は認知障害者などの視聴者にとってのメディアコンテンツの利用しやすさを強化するものである。これらのビデオ説明は、事前録画されたメディアコンテンツのオーディオトラック（例えば、せりふ）間の自然な途切れに挿入される。自然な途切れにおけるビデオ説明の挿入に関するいくつかのシナリオでは、対応する自然な途切れの期間内にビデオ説明が収まるように、対応するビデオ説明の１又は２以上の関連部分を削除し、又は自然な途切れの期間を増加させる編集が行われる。このようなシナリオでは、ビデオ説明の関連部分の削除又はメディアコンテンツのオーディオトラックの期間の増大が望ましくない場合もあり、視聴者のコンテンツ体験が不快で低品質なものになってしまう恐れがある。さらに、メディアコンテンツのせりふの自然な途切れにビデオ説明が挿入されるので、認知障害者は、自然な老化過程の一部としてよく理解することができず、従ってビデオ説明を理解できないことが多い。従って、視聴者（例えば、視覚障害者又は認知障害者）のメディアコンテンツ体験を改善するようにビデオ説明を効果的に制御できる強化された装置が必要とされている。

【発明の概要】

【発明が解決しようとする課題】

【０００４】

当業者には、説明したシステムと、本出願の残り部分において図面を参照しながら示す本開示のいくつかの態様とを比較することにより、従来の慣習的な手法のさらなる限界及び不利点が明らかになるであろう。

【課題を解決するための手段】

【０００５】

実質的に少なくとも１つの図に関連して図示及び／又は説明し、特許請求の範囲にさらに完全に示すような、シーン説明の再生制御のためのメディアレンダリング装置及び方法を提供する。

【０００６】

全体を通じて同じ要素を同じ参照符号によって示す添付図面を参照しながら本開示の以下の詳細な説明を検討することにより、本開示のこれらの及びその他の特徴及び利点を理解することができる。

【図面の簡単な説明】

【０００７】

【図１】本開示の実施形態による、シーン説明の再生制御のための例示的なネットワーク環境を示すブロック図である。

【図２】本開示の実施形態による、シーン説明の再生制御のための例示的なメディアレンダリング装置を示すブロック図である。

【図３Ａ】本開示の実施形態による、シーン説明の再生制御のための例示的なシナリオを図３Ｂと合わせて示す図である。

【図３Ｂ】本開示の実施形態による、シーン説明の再生制御のための例示的なシナリオを図３Ａと合わせて示す図である。

【図４】本開示の実施形態による、シーン説明の再生制御のための別の例示的なシナリオを示す図である。

【図５】本開示の実施形態による、シーン説明の再生制御のための例示的な動作を示す第１のフローチャートである。

【図６】本開示の実施形態による、シーン説明の再生制御のための例示的な動作を示す第２のフローチャートである。

【発明を実施するための形態】
【０００８】
開示する（視覚障害者又は認知障害者ユーザなどの）視聴者のメディアコンテンツ体験を強化するシーン説明の再生の動的制御のためのメディアレンダリング装置及び方法では、後述する実装を見出すことができる。本開示の例示的な態様は、一連の撮影シーンを含むことができるメディアコンテンツ（例えば、映画）を検索するように構成できるメディアレンダリング装置（例えば、テレビ）を提供する。メディアコンテンツは、ビデオ説明情報（例えば、一連の撮影シーンにおける撮影シーンを説明できるビデオ、筋書き又はシーン説明）と、ビデオ説明情報を再生するためのタイミング情報とを含むことができるテキスト情報をさらに含むことができる。タイミング情報は、ビデオ説明情報のテキスト表現又はオーディオ表現、或いはこれらの組み合わせを収めることができる空白又は途切れ（すなわち、メディアコンテンツのオーディオ部分の空白）に関する情報を含むことができる。メディアレンダリング装置は、ビデオ説明情報を再生するために撮影シーンのテキスト情報からタイミング情報を抽出することができる。メディアレンダリング装置は、ビデオ説明情報の再生を、第１の時間間隔（すなわち、撮影シーンの抽出されたタイミング情報によって示される第１の時間間隔）においてオーディオ表現、テキスト表現、又はテキスト表現及びオーディオ表現で制御するように構成することができる。
【０００９】
別の実施形態では、テキスト情報が、ビデオ説明情報を再生するための速度情報をさらに含むことができる。速度情報は、タイミング情報に対応するビデオ説明情報のオーディオ表現を再生するための再生速度に関する情報を含むことができる。メディアレンダリング装置は、ビデオ説明情報のオーディオ表現を再生するために撮影シーンのテキスト情報から速度情報を抽出することができる。メディアレンダリング装置は、抽出された速度情報に基づいて、第１の時間間隔（すなわち、抽出された撮影シーンのタイミング情報によって示される第１の時間間隔）においてビデオ説明情報のオーディオ表現の再生を制御するように構成することができる。
【００１０】
別の実施形態では、メディアレンダリング装置を、一連の撮影シーンのみを含むことができるメディアコンテンツと、一連の撮影シーンにおける撮影シーンを説明することはできるがタイミング情報及び速度情報を含まないビデオ説明情報とを検索するように構成することができる。メディアレンダリング装置は、撮影シーンにおけるオーディオ部分（例えば、せりふ）を再生するための時間間隔をそれぞれが示すことができる、撮影シーンの一連の第２の時間間隔を決定するように構成することができる。メディアレンダリング装置は、撮影シーンのビデオ説明情報（すなわち、シーン説明）のオーディオ表現をレンダリングする期間に対応できる第３の時間間隔を決定するようにさらに構成することができる。メディアレンダリング装置は、ビデオ説明情報のオーディオ表現を含めるために、一連の第２の時間間隔の合間の第１の時間間隔（すなわち、空白又は途切れ）を決定し、決定された一連の第２の時間間隔及び決定された第３の時間間隔に基づいて、含められるビデオ説明情報のオーディオ表現の再生速度をさらに制御するように構成することができる。
【００１１】
別の実施形態では、メディアレンダリング装置が、メディアレンダリング装置１０２に関連するディスプレイ装置上にビデオ説明情報を（例えば、テキストフォーマット又は表現で）直接レンダリングすることができる。ビデオ説明情報のテキストは、ディスプレイ装置上にレンダリングできる検索されたメディアコンテンツ上に、又は検索されたメディアコンテンツの外部にオーバーレイ表示することができる。別の実施形態では、ビデオ説明情報のテキストを、任意にクローズドキャプション情報（すなわち、メディアコンテンツのオーディオ部分又はせりふに関連するクローズドキャプション）と共に表示することができる。このことは、ビデオ説明情報が長く、ユーザがディスプレイ装置上にレンダリングされたビデオ説明情報を読むためにさらなる時間を必要とする場合に、メディアレンダリング装置のユーザがメディアコンテンツを手動で制御（一時停止及び再生）すること

ができるため有利である。

【００１２】

メディアコンテンツの途切れ／空白にシーン説明を含めるためにビデオ／シーン説明の関連部分を削除し、又は途切れ／空白の長さの期間を増加させることがある従来の解決策とは対照的に、開示するメディアレンダリング装置は、メディアコンテンツと共に検索できる、又は撮影シーンのせりふ間の検出された空白／途切れの期間に基づいて動的に決定できる速度に基づいて、ビデオ説明情報（すなわち、シーン又はビデオ説明）のオーディオ表現を再生することができる。メディアレンダリング装置は、撮影シーンにおいて識別される（単複の）自然な途切れ又は空白の期間と、メディアレンダリング装置に関連する規定の速度設定とに基づいて、ビデオ説明情報のオーディオ表現の再生速度を増加／減少させることができる。従って、シーン／ビデオ説明の全体的な再生品質が損なわれず、これによって視聴者（視覚障害者又は認知障害者）のコンテンツ体験をリアルタイムベースでさらに強化することができる。

【００１３】

さらに、ビデオ又はシーン説明をオーディオ形態で受け取ることができる従来の解決策と比べて、開示するメディアレンダリング装置は、シーン説明をテキストフォーマットで受け取り、又はテキストフォーマットでメディアコンテンツに含め、さらにシーン説明のテキスト情報をオーディオ表現に変換するように構成することができる。なお、任意に、ビデオ説明情報は、ディスプレイ装置上に直接レンダリングされるようにテキストとして保持する（すなわち、メディアコンテンツ上に、又はメディアコンテンツへの影響が大きい場合にはメディアコンテンツの外部にオーバーレイ表示する）こともできる。従って、シーン説明をテキストフォーマットで含め又は伝えることで、２つの装置間でビデオ説明をオーディオフォーマットで送信するのに必要とされる適切な帯域幅を節約することができる。従って、開示するメディアレンダリング装置は、ビデオ説明情報を含むオーディオトラックとビデオ説明情報を含まないオーディオトラックとを基本的に重複させる従来の解決策と比べて帯域幅を効率的に利用することができる。また、テキスト版のビデオ説明は、映画又はＴＶ番組内の特定のシーンを検索するための単語検索を可能にすることもできる。

【００１４】

図１は、本開示の実施形態による、シーン説明の再生制御のための例示的なネットワーク環境を示すブロック図である。図１にはネットワーク環境１００を示す。ネットワーク環境１００は、メディアレンダリング装置１０２、ディスプレイ装置１０４、オーディオレンダリング装置１０６、サーバ１０８、及び通信ネットワーク１１０を含むことができる。メディアレンダリング装置１０２は、通信ネットワーク１１０を介してディスプレイ装置１０４、オーディオレンダリング装置１０６、サーバ１０８に通信可能に結合することができる。メディアレンダリング装置１０２は、アンテナに接続された時にメディアコンテンツ１１２を受信できるように無線地上波チューナ（図示せず）と共に構成することができる。図１では、メディアレンダリング装置１０２及びディスプレイ装置１０４を２つの独立した装置として示しているが、いくつかの実施形態では、本開示の範囲から逸脱することなく、ディスプレイ装置１０４の機能全体をメディアレンダリング装置１０２に含めることもできる。

【００１５】

さらに、図１では、オーディオレンダリング装置１０６をメディアレンダリング装置１０２及び／又はディスプレイ装置１０４から分離して示しているが、本開示はこのように限定されるものではない。いくつかの実施形態では、本開示の範囲から逸脱することなく、オーディオレンダリング装置１０６をメディアレンダリング装置１０２及び／又はディスプレイ装置１０４に統合することもできる。図１には、一連の撮影シーン１１４、オーディオ部分１１６及びテキスト情報１１８を含むことができるメディアコンテンツ１１２をさらに示す。図１に示すように、一連の撮影シーン１１４は、第１の撮影シーン１１４Ａ、第２の撮影シーン１１４Ｂ、及び第Ｎの撮影シーン１１４Ｎを含むことができる。一

連の撮影シーン１１４の各々は、対応する撮影シーンを形成するように構築できる複数のショットを含むことができる。テキスト情報１１８は、ビデオ説明情報１１８Ａ及びタイミング情報１１８Ｂを含むこともできる。いくつかの実施形態では、テキスト情報が速度情報１１８Ｃを含むこともできる。ビデオ説明情報１１８Ａは、一連の撮影シーン１１４における少なくとも１つの撮影シーン（例えば、第１の撮影シーン１１４Ａ）に関する説明を含むことができる。いくつかの実施形態では、複数のショットを含む一連の撮影シーン１１４の各々が、オーディオ部分１１６に関連する１又は２以上の画像フレーム又は部分をさらに含むことができる。さらに、メディアレンダリング装置１０２に関連することができるユーザ１２０も示す。例えば、ユーザ１２０は、メディアコンテンツ１１２の視聴者であることができ、視覚障害又は認知障害視聴者とすることができる。

【００１６】

メディアレンダリング装置１０２は、（サーバ１０８などの）リモートソース又はメディアレンダリング装置１０２のメモリ（すなわち、図２のメモリ２０４）からメディアコンテンツ１１２を検索するように構成できる好適なロジック、回路、インターフェイス及び／又はコードを含むことができる。いくつかの実施形態では、地上波チューナを利用して無線でメディアコンテンツ１１２を検索することができる。いくつかのシナリオでは、高度テレビシステム委員会（ＡＴＳＣ）又はＡＴＳＣ　３．０標準を使用して、メディアコンテンツ１１２をデジタルで受信することができる。

【００１７】

メディアコンテンツ１１２は、ビデオ説明情報１１８Ａ、タイミング情報１１８Ｂ及び速度情報１１８Ｃを含むことができるテキスト情報１１８を含むことができる。ビデオ説明情報１１８Ａは、一連の撮影シーン１１４のうちの（第１の撮影シーン１１４Ａなどの）撮影シーンを説明することができる。メディアレンダリング装置１０２は、第１の撮影シーン１１４Ａのテキスト情報１１８からタイミング情報１１８Ｂを抽出するように構成することができる。タイミング情報１１８Ｂは、メディアレンダリング装置１０２がビデオ説明情報を再生するために使用することができる。いくつかの実施形態では、メディアレンダリング装置が、ビデオ説明情報を再生するためにタイミング情報１１８Ｂと共に速度情報１１８Ｃを使用することもできる。メディアレンダリング装置１０２は、ビデオ説明情報の再生を、撮影シーンの抽出されたタイミング情報によって示される第１の時間間隔においてテキスト表現、オーディオ表現、又はテキスト表現及びオーディオ表現の両方で制御するようにさらに構成することができる。他のいくつかの実施形態では、メディアレンダリング装置１０２を、抽出された速度情報１１８Ｃに基づいて、撮影シーンの抽出されたタイミング情報によって示される第１の時間間隔においてビデオ説明情報のオーディオ表現の再生を制御するようにさらに構成することができる。

【００１８】

別の実施形態では、メディアレンダリング装置１０２が、（サーバ１０８などの）リモートソース又はメディアレンダリング装置１０２のメモリ（すなわち、図２のメモリ２０４）からメディアコンテンツ１１２を検索することができる。メディアコンテンツは、一連の撮影シーン１１４の（第１の撮影シーン１１４Ａなどの）撮影シーンを説明できるビデオ説明情報１１８Ａを含むことができる。メディアレンダリング装置１０２は、第１の撮影シーン１１４Ａのオーディオ部分１１６を再生するための一連の第２の時間間隔を決定し、ビデオ説明情報１１８Ａのオーディオ表現を再生するための第３の時間間隔を決定するように構成することができる。メディアレンダリング装置１０２は、決定された一連の第２の時間間隔及び第３の時間間隔に基づいて、ビデオ説明情報１１８Ａのオーディオ表現の再生速度を決定するようにさらに構成することができる。メディアレンダリング装置１０２の例としては、以下に限定するわけではないが、デジタルメディアプレーヤ（ＤＭＰ）、スマートテレビメディアプレーヤ、オーバーザトップ（ＯＴＴ）プレーヤ、デジタルメディアストリーマ、メディアエクステンダ／レギュレータ、デジタルメディアハブ、メディアコンテンツコントローラ、テレビ、コンピュータワークステーション、メインフレームコンピュータ、ハンドヘルドコンピュータ、携帯電話機、スマートフォン、セル

ラー電話機、スマート家電、携帯情報端末（ＰＤＡ）、スマートスピーカ、スマートメガネ、サウンドシステム、ヘッドマウント装置（ＨＭＤ）、ヘッドセット、スマートヘッドホン、及び／又はオーディオ－ビデオレンダリング能力を有するその他のコンピュータ装置を挙げることができる。

【００１９】

ディスプレイ装置１０４は、検索されたメディアコンテンツ１１２内に存在する一連の撮影シーン１１４を表示するように構成できる好適なロジック、回路及びインターフェイスを含むことができる。ディスプレイ装置１０４は、ビデオ説明情報１１８Ａをテキストフォーマットで表示するようにさらに構成することができる。ディスプレイ装置１０４は、ユーザがディスプレイ装置１０４を介してユーザ入力を提供することを可能にするタッチ画面とすることができる。タッチ画面は、抵抗膜式タッチ画面、静電容量式タッチ画面、又は感熱式タッチ画面のうちの少なくとも１つとすることができる。ディスプレイ装置１０４は、以下に限定するわけではないが、液晶ディスプレイ（ＬＣＤ）ディスプレイ、発光ダイオード（ＬＥＤ）ディスプレイ、プラズマディスプレイ、又は有機ＬＥＤ（ＯＬＥＤ）ディスプレイ技術のうちの少なくとも１つ、或いはその他のディスプレイ装置などの複数の既知の技術を通じて実現することができる。ある実施形態によれば、ディスプレイ装置１０４は、ヘッドマウント装置（ＨＭＤ）のディスプレイ画面、スマートメガネ装置、シースルーディスプレイ、投影型ディスプレイ、エレクトロクロミックディスプレイ、又は透明ディスプレイを意味することができる。

【００２０】

オーディオレンダリング装置１０６は、ビデオ説明情報１１８Ａ（すなわち、シーン又はビデオ説明）のオーディオ表現を再生又はプレイバックするように構成できる好適なロジック、回路及びインターフェイスを含むことができる。オーディオレンダリング装置１０６は、第１の撮影シーン１１４Ａ又は一連の撮影シーン１１４のオーディオ部分１１６（例えば、せりふ）を再生するようにさらに構成することができる。オーディオレンダリング装置１０６の例としては、以下に限定するわけではないが、ラウドスピーカ、壁埋め込み型／天井取り付け型スピーカ、サウンドバー、ウーファ又はサブウーファ、サウンドカード、ヘッドフォン、ヘッドセット、ワイヤレススピーカ、及び／又はオーディオ再生能力を有するその他のコンピュータ装置を挙げることができる。

【００２１】

サーバ１０８は、メディアコンテンツ１１２を記憶するように構成できる好適なロジック、回路、インターフェイス及びコードを含むことができる。サーバ１０８は、メディアレンダリング装置１０２から、サーバ１０８に記憶されているメディアコンテンツ１１２を検索するための要求を受け取ることができる。いくつかの実施形態では、サーバ１０８を、ビデオ説明情報１１８Ａ（すなわち、シーン説明）のオーディオ表現の最大再生速度を示すことができる規定の速度設定を記憶するように構成することができる。サーバ１０８は、ウェブアプリケーション、クラウドアプリケーション、ＨＴＴＰリクエスト、リポジトリ操作及びファイル転送などを通じて動作を実行できるクラウドサーバとして実装することができる。サーバ１０８の他の例としては、以下に限定するわけではないが、データベースサーバ、ファイルサーバ、ウェブサーバ、メディアサーバ、アプリケーションサーバ、メインフレームサーバ、クラウドサーバ、又はその他のタイプのサーバを挙げることができる。１又は２以上の実施形態では、サーバ１０８を、当業者に周知の複数の技術を使用することによって複数の分散型クラウドベースリソースとして実装することができる。当業者であれば、本開示の範囲は、サーバ１０８及びメディアレンダリング装置１０２を独立エンティティとして実装することに限定されるものではないと理解するであろう。いくつかの実施形態では、本開示の範囲から逸脱することなく、サーバ１０８の機能を全体的に又は少なくとも部分的にメディアレンダリング装置１０２に組み込むこともできる。

【００２２】

通信ネットワーク１１０は、メディアレンダリング装置１０２、ディスプレイ装置１０

４、オーディオレンダリング装置１０６及びサーバ１０８が互いに通信できるようにする通信媒体を含むことができる。通信ネットワーク１１０は、有線通信ネットワーク又は無線通信ネットワークとすることができる。通信ネットワーク１１０の例としては、以下に限定するわけではないが、インターネット、クラウドネットワーク、ワイヤレスフィデリティ（Ｗｉ－Ｆｉ）ネットワーク、パーソナルエリアネットワーク（ＰＡＮ）、ローカルエリアネットワーク（ＬＡＮ）、又はメトロポリタンエリアネットワーク（ＭＡＮ）を挙げることができる。ネットワーク環境１００内の様々な装置は、様々な有線及び無線通信プロトコルに従って通信ネットワーク１１０に接続するように構成することができる。このような有線及び無線通信プロトコルの例としては、以下に限定するわけではないが、伝送制御プロトコル及びインターネットプロトコル（ＴＣＰ／ＩＰ）、ユーザデータグラムプロトコル（ＵＤＰ）、ハイパーテキスト転送プロトコル（ＨＴＴＰ）、ファイル転送プロトコル（ＦＴＰ）、ＺｉｇＢｅｅ、ＥＤＧＥ、ＩＥＥＥ８０２．１１、ライトフィデリティ（Ｌｉ－Ｆｉ）、８０２．１６、ＩＥＥＥ８０２．１１ｓ、ＩＥＥＥ８０２．１１ｇ、マルチホップ通信、無線アクセスポイント（ＡＰ）、装置間通信、セルラー通信プロトコル、Ｂｌｕｅｔｏｏｔｈ（ＢＴ）通信プロトコルを挙げることができる。

【００２３】

動作時には、開示するメディアレンダリング装置１０２が、ユーザ１２０からメディアコンテンツ１１２を再生するための要求を受け取ることができる。メディアコンテンツ１１２の例としては、以下に限定するわけではないが、ビデオクリップ、映画、広告、オーディオ－ビデオコンテンツ、ゲームコンテンツ、又はスライドショークリップを挙げることができる。メディアレンダリング装置１０２は、この要求に基づいて、（サーバ１０８などの）リモートソース又はメディアレンダリング装置１０２の（図２のメモリ２０４などの）メモリからメディアコンテンツ１１２を検索することができる。メディアコンテンツ１１２は、一連の撮影シーン１１４、オーディオ部分１１６、及びテキスト情報１１８を含むことができる。テキスト情報１１８は、一連の撮影シーン１１４のうちの撮影シーン（例えば、第１の撮影シーン１１４Ａ）を説明することができる、テキストフォーマットであることができるビデオ説明情報１１８Ａを含むことができる。いくつかの実施形態では、ビデオ説明情報１１８Ａが、メディアコンテンツ１１２内に存在する一連の撮影シーン１１４の各々を説明することができる。ある実施形態では、メディアレンダリング装置１０２を、第１の撮影シーン１１４Ａの（例えば、テキストフォーマットの）ビデオ説明情報１１８Ａをビデオ説明情報１１８Ａのオーディオ表現に変換するようにさらに構成することができる。テキスト情報１１８は、タイミング情報１１８Ｂを含むこともできる。タイミング情報１１８Ｂは、ビデオ説明情報１１８Ａのオーディオ表現を収めて再生できる第１の時間間隔を示すことができる。別の実施形態では、テキスト情報１１８が速度情報１１８Ｃをさらに含むことができる。速度情報１１８Ｃは、タイミング情報１１８Ｂによって示される（第１の時間間隔などの）特定の時間間隔中にビデオ説明情報１１８Ａのオーディオ表現を再生する再生速度を示すことができる。メディアレンダリング装置１０２は、第１の撮影シーン１１４Ａのテキスト情報１１８からタイミング情報１１８Ｂを抽出するようにさらに構成することができる。メディアレンダリング装置１０２は、ビデオ説明情報１１８Ａの再生を、一連の撮影シーン１１４の第１の撮影シーン１１４Ａの抽出されたタイミング情報１１８Ｂによって示される第１の時間間隔においてテキスト表現、オーディオ表現、又はテキスト表現及びオーディオ表現で制御するようにさらに構成することができる。他のいくつかの実施形態では、メディアレンダリング装置１０２を、速度情報１１８Ｃを抽出するようにさらに構成することができる。このような事例では、メディアレンダリング装置１０２を、抽出された速度情報１１８Ｃに基づいて、一連の撮影シーン１１４のうちの第１の撮影シーン１１４Ａの抽出されたタイミング情報１１８Ｂによって示される第１の時間間隔においてビデオ説明情報１１８Ａのオーディオ表現の再生を制御するようにさらに構成することができる。

【００２４】

別の実施形態では、メディアレンダリング装置１０２が、一連の撮影シーン１１４にお

ける第１の撮影シーン１１４Ａのオーディオ部分１１６（すなわち、せりふ）を再生するための時間間隔をそれぞれが示すことができる、第１の撮影シーン１１４Ａの一連の第２の時間間隔を決定することができる。メディアレンダリング装置１０２は、第１の撮影シーン１１４Ａのビデオ説明情報１１８Ａのオーディオ表現を再生するために必要な第３の時間間隔を決定するようにさらに構成することができる。第３の時間間隔は、第１の撮影シーン１１４Ａのビデオ説明情報１１８Ａのオーディオ表現を再生するのにかかる時間又はそのために必要な期間に対応することができる。一連の第２の時間間隔及び第３の時間間隔の詳細については、例えば図４で説明する。

【００２５】

メディアレンダリング装置１０２は、ビデオ説明情報１１８Ａのオーディオ表現を再生する速度を決定するようにさらに構成することができる。決定される速度は、例えば第１の撮影シーン１１４Ａの再生中にユーザ１２０のためにビデオ説明情報１１８Ａのオーディオ表現を再生できる速度とすることができる。ビデオ説明情報１１８Ａのオーディオ表現の再生速度は、決定された一連の第２の時間間隔及び決定された第３の時間間隔に基づいて決定することができる。いくつかの実施形態では、決定される速度が、ビデオ説明情報１１８Ａのオーディオ表現の実際の再生速度よりも低いことができる。他のいくつかの実施形態では、決定される速度が、ビデオ説明情報１１８Ａのオーディオ表現の実際の再生速度よりも高いことができる。決定された一連の第２の時間間隔及び決定された第３の時間間隔に基づくビデオ説明情報１１８Ａのオーディオ表現の再生速度の決定の詳細については、例えば図４で説明する。

【００２６】

メディアレンダリング装置１０２は、決定された速度に基づいて、ビデオ説明情報１１８Ａのオーディオ表現の再生を制御するようにさらに構成することができる。ビデオ説明情報１１８Ａのオーディオ表現は、第１の時間間隔（例えば、第１の撮影シーン１１４Ａのせりふ間の空白）において再生することができる。第１の時間間隔は、一連の第２の時間間隔とは異なることができる。いくつかの実施形態では、第１の時間間隔を、第１の撮影シーン１１４Ａの第１のせりふと第２のせりふとの間の間隔（すなわち、空白）とすることができる。第１のせりふは、第１の撮影シーン１１４Ａのあるショット（例えば、第１のショット）の最後の単語に対応することができ、第２のせりふは、第１の撮影シーン１１４Ａの次のショット（例えば、第２のショット）の最初の単語に対応することができる。第１のショット及び第２のショットは、第１の撮影シーン１１４Ａの連続するショットとすることができる。別の実施形態では、第１の時間間隔を、第１の撮影シーン１１４Ａの開始と第１の撮影シーン１１４Ａの第１のせりふとの間の間隔（すなわち、空白）とすることができる。ある実施形態では、第１の時間間隔（すなわち、空白）が第３の時間間隔よりも短い場合、メディアレンダリング装置１０２が、ビデオ説明情報１１８Ａのオーディオ表現の再生速度をビデオ説明情報１１８Ａのオーディオ表現の実際の再生速度よりも高くなるように決定することができる。別の実施形態では、第１の時間間隔（すなわち、空白）が第３の時間間隔よりも長い場合、メディアレンダリング装置１０２が、ビデオ説明情報１１８Ａのオーディオ表現の再生速度をオーディオ表現の実際の再生速度よりも低くなるように決定することができる。ビデオ説明情報１１８Ａのオーディオ表現（すなわち、第１の撮影シーン１１４Ａの説明）の再生速度を増減することで、（説明内容の短縮のような）著しい修正を伴わずにシーン説明を再生することができ、視覚障害又は認知障害視聴者などのユーザ１２０のためにシーン／ビデオ説明の品質をさらに維持することができる。

【００２７】

図１には、本開示の範囲から逸脱することなく修正、追加又は省略を行うことができる。例えば、ネットワーク環境１００は、本開示において図示し説明する要素よりも多くの又は少ない要素を含むことができる。例えば、いくつかの実施形態では、ネットワーク環境１００が、メディアレンダリング装置１０２を含んでディスプレイ装置１０４を含まないこともできる。また、いくつかの実施形態では、本開示の範囲から逸脱することなく、

各オーディオレンダリング装置１０６の機能をディスプレイ装置１０４に組み込むこともできる。

【００２８】

図２は、本開示の実施形態による、シーン説明の再生制御のための例示的なメディアレンダリング装置を示すブロック図である。図２の説明は、図１の要素に関連して行う。図２には、メディアレンダリング装置１０２のブロック図２００を示す。メディアレンダリング装置１０２は、シーン又はビデオ説明の再生を制御する動作を実行できる回路２０２を含むことができる。メディアレンダリング装置１０２は、メモリ２０４、入力／出力（Ｉ／Ｏ）装置２０６、テキスト－スピーチ変換器２０８、ネットワークインターフェイス２１０、ディスプレイ装置１０４、及びオーディオレンダリング装置１０６をさらに含むことができる。メモリ２０４は、機械学習（ＭＬ）モデル２１２を含むことができる。回路２０２は、メモリ２０４、Ｉ／Ｏ装置２０６、テキスト－スピーチ変換器２０８、ネットワークインターフェイス２１０、ディスプレイ装置１０４及びオーディオレンダリング装置１０６に通信可能に結合することができる。

【００２９】

回路２０２は、メディアレンダリング装置１０２によって実行される異なる動作に関連するプログラム命令を実行するように構成できる好適なロジック、回路及びインターフェイスを含むことができる。例えば、これらの動作の一部は、メディアコンテンツ１１２の検索、タイミング情報１１８Ｂ及び／又は速度情報１１８Ｃの抽出、及び抽出されたタイミング情報１１８Ｂ及び速度情報１１８Ｃに基づく第１の時間間隔におけるビデオ説明情報１１８Ａのテキスト表現又はオーディオ表現又はテキスト表現及びオーディオ表現の再生を含むことができる。回路２０２は、独立したプロセッサとして実装できる１又は２以上の特殊処理ユニットを含むことができる。ある実施形態では、１又は２以上の特殊処理ユニットを、１又は２以上の特殊処理ユニットの機能をまとめて実行する統合プロセッサ又はプロセッサ群として実装することができる。回路２０２は、当業で周知の複数のプロセッサ技術に基づいて実装することができる。回路２０２の実装例は、Ｘ８６ベースのプロセッサ、グラフィックプロセッシングユニット（ＧＰＵ）、縮小命令セットコンピューティング（ＲＩＳＣ）プロセッサ、特定用途向け集積回路（ＡＳＩＣ）プロセッサ、複合命令セットコンピューティング（ＣＩＳＣ）プロセッサ、マイクロコントローラ、中央処理装置（ＣＰＵ）、及び／又はその他の制御回路とすることができる。

【００３０】

メモリ２０４は、回路２０２によって実行される命令を記憶するように構成できる好適なロジック、回路、インターフェイス及び／又はコードを含むことができる。メモリ２０４は、メディアコンテンツ１１２、テキスト情報１１８、及びビデオ説明情報１１８Ａのオーディオ表現の最大再生速度を示すことができる規定の速度設定を記憶するように構成することができる。メモリ２０４は、（第１の撮影シーン１１４Ａなどの）撮影シーンのコンテキスト情報に基づいてオーディオ特性を決定するように構成できる訓練済み機械学習（ＭＬ）モデル２１２を記憶するようにさらに構成することができる。ＭＬモデル２１２の詳細な機能については、例えば図４で説明する。メモリ２０４は、ユーザのプロファイル情報を記憶するようにさらに構成することができる。メモリ２０４の実装例としては、以下に限定するわけではないが、ランダムアクセスメモリ（ＲＡＭ）、リードオンリメモリ（ＲＯＭ）、電気的に消去可能なプログラマブルリードオンリメモリ（ＥＥＰＲＯＭ）、ハードディスクドライブ（ＨＤＤ）、固体ドライブ（ＳＳＤ）、ＣＰＵキャッシュ、及び／又はセキュアデジタル（ＳＤ）カードなどを挙げることができる。

【００３１】

Ｉ／Ｏ装置２０６は、入力を受け取り、受け取った入力に基づいて出力を提供するように構成できる好適なロジック、回路及びインターフェイスを含むことができる。Ｉ／Ｏ装置２０６は、撮影シーンの説明に対応する第１のユーザ入力を受け取るように構成することができる。Ｉ／Ｏ装置は、ディスプレイ装置１０４及びオーディオレンダリング装置１０６を介してメディアコンテンツ１１２の再生を制御するようにさらに構成することがで

きる。Ｉ／Ｏ装置は、回路２０２と通信するように構成できる様々な入力及び出力装置を含むことができる。Ｉ／Ｏ装置２０６の例としては、以下に限定するわけではないが、ディスプレイ装置１０４、オーディオレンダリング装置１０６、タッチ画面、キーボード、マウス、ジョイスティック、及びマイクを挙げることができる。

【００３２】

テキスト－スピーチ変換器２０８は、一連の撮影シーン１１４のうちの少なくとも第１の撮影シーン１１４Ａを説明するビデオ説明情報１１８Ａをオーディオレンダリング装置１０６による再生に適合できるオーディオフォーマットに変換するように構成できる好適なロジック、回路、インターフェイス及び／又はコードを含むことができる。本明細書では、変換されたオーディオをビデオ説明情報１１８Ａのオーディオ表現と呼ぶことができ、オーディオレンダリング装置１０６上でレンダリングすることができる。テキスト－スピーチ変換器２０８は、当業で周知の数多くのプロセッサ技術に基づいて実装することができる。プロセッサ技術の例としては、以下に限定するわけではないが、中央処理装置（ＣＰＵ）、ｘ８６ベースのプロセッサ、縮小命令セットコンピューティング（ＲＩＳＣ）プロセッサ、特定用途向け集積回路（ＡＳＩＣ）プロセッサ、複合命令セットコンピューティング（ＣＩＳＣ）プロセッサ、及びその他のプロセッサを挙げることができる。

【００３３】

ネットワークインターフェイス２１０は、通信ネットワーク１１０を介して回路２０２とサーバ１０８との間の通信を容易にするように構成できる好適なロジック、回路及びインターフェイスを含むことができる。ネットワークインターフェイス２１０は、メディアレンダリング装置１０２と通信ネットワーク１１０との有線又は無線通信をサポートする様々な既知の技術を使用して実装することができる。ネットワークインターフェイス２１０は、以下に限定するわけではないが、アンテナ、無線周波数（ＲＦ）トランシーバ、１又は２以上の増幅器、チューナ、１又は２以上の発振器、デジタルシグナルプロセッサ、コーダ－デコーダ（ＣＯＤＥＣ）チップセット、加入者ＩＤモジュール（ＳＩＭ）カード、又はローカルバッファ回路を含むことができる。ネットワークインターフェイス２１０は、インターネット、イントラネットなどのネットワーク、又はセルラー電話ネットワーク、無線ローカルエリアネットワーク（ＬＡＮ）及びメトロポリタンエリアネットワーク（ＭＡＮ）などの無線ネットワークと無線通信を介して通信するように構成することができる。無線通信は、グローバルシステムフォーモバイルコミュニケーションズ（ＧＳＭ）、拡張データＧＳＭ環境（ＥＤＧＥ）、広帯域符号分割多重アクセス（Ｗ－ＣＤＭＡ）、ロングタームエボリューション（ＬＴＥ）、符号分割多重アクセス（ＣＤＭＡ）、時分割多重アクセス（ＴＤＭＡ）、Ｂｌｕｅｔｏｏｔｈ、（ＩＥＥＥ８０２．１１ａ、ＩＥＥＥ８０２．１１ｂ、ＩＥＥＥ８０２．１１ｇ又はＩＥＥＥ８０２．１１ｎなどの）ワイヤレスフィデリティ（ＷｉＦｉ）、ボイスオーバーインターネットプロトコル（ＶｏＩＰ）、ライトフィデリティ（Ｌｉ－Ｆｉ）、ワールドワイド・インターオペラビリティ・フォー・マイクロウェーブ・アクセス（Ｗｉ－ＭＡＸ）、電子メール用プロトコル、インスタントメッセージ、及びショートメッセージサービス（ＳＭＳ）などの複数の通信標準、プロトコル及び技術のうちの１つ又は２つ以上を使用するように構成することができる。

【００３４】

図３Ａは、本開示の実施形態による、シーン説明の再生制御のための例示的なシナリオを示す図である。図３Ａの説明は、図１及び図２の要素に関連して行う。図３Ａには例示的なシナリオ３００を示す。例示的なシナリオ３００には、メディアレンダリング装置３０２（すなわち、メディアレンダリング装置１０２と同様のもの）を示す。図３Ａには、メディアレンダリング装置３０２に関連するディスプレイ装置１０４及びオーディオレンダリング装置１０６をさらに示す。メディアレンダリング装置３０２は、ディスプレイ装置１０４及びオーディオレンダリング装置１０６を、メディアコンテンツをレンダリングするように制御することができる。メディアコンテンツの例としては、以下に限定するわけではないが、ビデオクリップ、映画、オーディオビデオコンテンツ、ゲームコンテンツ、広告、又はスライドショークリップを挙げることができる。メディアコンテンツは、（

図３Ａに示すような）ディスプレイ装置１０４上に表示された撮影シーン３０４を含むことができる（図１の一連の撮影シーン１１４などの）一連の撮影シーンを含むことができる。

【００３５】

なお、図３Ａに示す撮影シーン３０４は車のシーンの一例として提示するものにすぎない。本開示は、（以下に限定するわけではないが、アクションシーン、ドラマシーン、ロマンチックなシーン、感情的なシーン、ダンスシーン、音楽シーン、ホラーシーン、又はラブシーンなどの）他のタイプの撮影シーンにも適用可能である。他のタイプの撮影シーンの説明については、簡潔にするために本開示からは省略する。さらに、図３Ａに示すディスプレイ装置１０４はテレビの一例として提示するものにすぎない。本開示は、例えば図１で説明したような他のタイプのディスプレイ装置にも適用可能である。このような他のタイプのディスプレイ装置の説明については、簡潔にするために本開示からは省略する。さらに、図３Ａに示すオーディオレンダリング装置１０６はスピーカの一例として提示するものにすぎない。本開示は、例えば図１で説明したような他のタイプのオーディオレンダリング装置にも適用可能である。このような他のタイプのオーディオレンダリング装置の説明については、簡潔にするために本開示からは省略している。

【００３６】

ある実施形態では、メディアコンテンツの一連の撮影シーンの各々が、オーディオ部分、画像部分、及びテキスト情報３０６を含むことができる。オーディオ部分は、オーディオレンダリング装置１０６を介してレンダリングできるオーディオフォーマットでの、メディアコンテンツの一連の撮影シーンの各々の１又は２以上のせりふを含むことができる。各シーンの画像部分は、ディスプレイ装置１０４上にレンダリングできる１又は２以上の画像フレームを含むことができる。

【００３７】

テキスト情報３０６は、ビデオ説明情報３０６Ａ、タイミング情報、及び／又は速度情報をさらに含むことができる。ビデオ説明情報３０６Ａは、一連の撮影シーンのうちの（撮影シーン３０４又は第１の撮影シーン１１４Ａなどの）撮影シーンを説明することができ、撮影シーン３０４はディスプレイ装置１０４上に表示することができる。いくつかの実施形態では、ビデオ説明情報３０６Ａが、一連の撮影シーンの各撮影シーンを説明することができる。ビデオ説明情報３０６Ａ又は撮影シーンの説明の例としては、以下に限定するわけではないが、撮影シーン内の１又は２以上の装飾品の説明、シーン内の照明条件の説明、撮影シーン内の場所の説明、撮影シーン内のカメラモーションの説明、撮影シーン内の背景情報の説明、撮影シーン内の環境条件の説明、撮影シーン内のショット推移の説明、撮影シーンに含まれるテキストの説明、撮影シーンに描かれるキャラクタの説明、撮影シーンに描かれるキャラクタの態度／感情の説明、撮影シーンに描かれるキャラクタ間の空間的関係の説明、撮影シーンに描かれるキャラクタの身体的属性の説明、撮影シーンに描かれるキャラクタの肉体的表現の説明、撮影シーンに描かれるキャラクタの表情の説明、撮影シーンに描かれるキャラクタの動きの説明、撮影シーンに描かれるキャラクタの職業又は役割の説明、撮影シーンに描かれるキャラクタの服装の説明などを挙げることができる。

【００３８】

ある実施形態によれば、回路２０２は、メディアレンダリング装置３０２のメモリ２０４から、一連の撮影シーン及びテキスト情報３０６を含むメディアコンテンツを検索するように構成することができる。いくつかの実施形態では、回路２０２を、メディアコンテンツを求める要求をサーバ１０８に送信するように構成することができる。送信された要求に基づいて、サーバ１０８から（一連の撮影シーン及びテキスト情報３０６を含むことができる）メディアコンテンツを受け取ることができる。テキスト情報３０６内に存在するビデオ説明情報３０６Ａは（オーディオフォーマットではなく）テキストフォーマットで受け取ることができ、これによりサーバ１０８とメディアレンダリング装置３０２との間におけるテキスト情報３０６の通信中の帯域幅をさらに節約することができる。テキス

トフォーマットでのテキスト情報３０６は、メモリ２０４又はサーバ１０８の記憶スペースをさらに節約することができる。ある実施形態では、メディアコンテンツの再生前に、メディアコンテンツから一連の撮影シーン及びテキスト情報３０６の各々を検索することができる。他のいくつかの実施形態では、回路２０２が、特定の撮影シーン（例えば、現在の撮影シーン）の再生時に、さらなる処理（例えば、次の撮影シーンのビデオ説明情報３０６Ａをオーディオ再生するためのタイミング情報及び速度情報の抽出又は速度の計算）のために次の撮影シーンのテキスト情報３０６を検索することができる。

【００３９】

ある実施形態では、撮影シーン３０４が、第１のショット及び第２のショットなどの複数のショットを含むことができる。撮影シーン３０４は、複数の生物オブジェクト（ａｎｉｍａｔｅｄ　ｏｂｊｅｃｔｓ）及び無生物オブジェクト（ｉｎ－ａｎｉｍａｔｅｄ　ｏｂｊｅｃｔｓ）を含むことができる。例えば、図３Ａでは、生物オブジェクトが、以下に限定するわけではないが、（例えば、「Ｇｅｏｒｇｅ」という名前の）第１の人物３０８及び（例えば、「Ｊａｃｋ」という名前の）第２の人物３１０を含むことができる。図３Ａに示す無生物オブジェクトは、限定するわけではないが車３１２を含むことができる。図３Ａにはタイムライン３１４をさらに示す。タイムライン３１４は、撮影シーン３０４を再生できる（例えば、秒、分又は時間単位の）時間間隔を示すことができる。タイムライン３１４に示される合計時間は、撮影シーン３０４に関連する全ての画像フレーム及び／又はオーディオフレームをレンダリングするための再生時間とすることができる。

【００４０】

図３Ａに示すように、撮影シーン３０４は、撮影シーン３０４のオーディオ部分３１６として第１のオーディオ部分３１６Ａ及び第２のオーディオ部分３１６Ｂを含むことができる。第１のオーディオ部分３１６Ａ及び第２のオーディオ部分３１６Ｂの各々は、撮影シーン３０４に取り込まれた第１の人物３０８及び／又は第２の人物３１０に対応する（図３Ａに示す「Ｇｅｏｒｇｅ：おいＪａｃｋ、どこに向かってるんだ？（Ｈｅｙ　Ｊａｃｋ！　Ｗｈｅｒｅ　ａｒｅ　ｙｏｕ　ｈｅａｄｉｎｇ）」及び「Ｊａｃｋ：仕事だよ（Ｉ　ａｍ　ｇｏｉｎｇ　ｔｏ　ｗｏｒｋ）」などの）せりふを含むことができる。一例として、図３Ａに示すように、撮影シーン３０４に関連するタイムライン３１４には、時点t_0～t_1に、第１の途切れ３１８Ａに対応できる自然な途切れが存在することができる。時点t_1～t_2には、第１の人物３０８が、せりふ又は第１のオーディオ部分３１６Ａとして「おいＪａｃｋ、どこに向かってるんだ？」と発話することができる。さらに、時点t_2～t_3には、第２の途切れ３１８Ｂに対応できる別の自然な途切れが存在することができる。時点t_3～t_4には、第２の人物３１０が、せりふ又は第２のオーディオ部分３１６Ｂとして、例えば「仕事だよ」というせりふで第１の人物３０８に返答することができる。時点t_4～t_5には、第３の途切れ３１８Ｃに対応できる別の自然な途切れが存在することができる。

【００４１】

一例として、ビデオ説明情報３０６Ａ、及びタイミング情報を含むことができる検索されたテキスト情報３０６を以下の表１に示す。

S. No	ビデオ説明	時間間隔
１.	Ｊａｃｋが車を運転中、ＧｅｏｒｇｅがＪａｃｋを見ている	t_2～t_3
２.	ビデオ説明１	t_A～t_B
３.	ビデオ説明２	t_C～t_D

表１：テキスト情報

【００４２】

なお、テキスト情報３０６内の行数は一例として提示するものにすぎない。テキスト情報３０６は、撮影シーン３０４に含まれるビデオ説明の数に基づいてこれよりも多くの又は少ない数の行を含むことができる。

【００４３】
回路２０２は、撮影シーン３０４のテキスト情報３０６からタイミング情報を抽出するようにさらに構成することができる。タイミング情報は、テキスト情報３０６のビデオ説明情報３０６Ａを再生するために抽出することができる。タイミング情報は、ビデオ説明情報３０６Ａのテキスト表現又はオーディオ表現、或いはテキスト表現及びオーディオ表現の両方を再生のために収めることができる、タイムライン３１４内の第１の時間間隔（例えば、第２の途切れ３１８Ｂとしての時間間隔t_2～t_3）を示すことができる。
【００４４】
別の実施形態では、回路２０２を、撮影シーン３０４のテキスト情報３０６から速度情報を抽出するようにさらに構成することができる。タイミング情報と同様に、速度情報も、テキスト情報３０６のビデオ説明情報３０６Ａを再生するために抽出することができる。速度情報は、タイミング情報１１８Ｂによって示される第１の時間間隔（すなわち、第２の途切れ３１８Ｂ）中にビデオ説明情報３０６Ａのオーディオ表現を再生する再生速度を示すことができる。一例として、ビデオ説明情報３０６Ａ、タイミング情報及び速度情報を含むことができる検索されたテキスト情報３０６を以下の表２に示す。

S. No	ビデオ説明	時間間隔	再生速度
1.	Ｊａｃｋが車を運転中、Ｇｅｏｒｇｅがｊａｃｋを見ている	t_2～t_3	1.6倍
2.	ビデオ説明1	t_A～t_B	0.5倍
3.	ビデオ説明2	t_C～t_D	2.0倍

表２：テキスト情報

【００４５】
なお、テキスト情報３０６内の行数は一例として提示するものにすぎない。テキスト情報３０６は、撮影シーン３０４に含まれるビデオ説明の数に基づいてこれよりも多くの又は少ない数の行を含むことができる。
【００４６】
テキスト表現の場合には、回路２０２を、撮影シーン３０４の抽出されたタイミング情報によって示される第１の時間間隔（すなわち、第２の途切れ３１８Ｂ）において（テキスト情報３０６内に存在する）ビデオ説明情報３０６Ａをディスプレイ装置１０４上にレンダリングするように構成することができる。ビデオ説明情報３０６Ａのテキスト再生に関する詳細については、例えば図３Ｂで説明する。
【００４７】
オーディオ表現の場合には、回路２０２を、撮影シーン３０４のテキスト情報３０６内に存在する検索されたビデオ説明情報３０６Ａをビデオ説明情報３０６Ａのオーディオ表現に変換するようにテキスト－スピーチ変換器２０８を制御するようさらに構成することができる。回路２０２は、撮影シーン３０４の抽出されたタイミング情報によって示される第１の時間間隔（すなわち、第２の途切れ３１８Ｂ）においてビデオ説明情報３０６Ａのオーディオ表現の再生を制御することができる。ビデオ説明情報３０６Ａのオーディオ表現の再生は、抽出された速度情報に基づくことができる。
【００４８】
テキスト表現及びオーディオ表現の両方の場合には、オーディオレンダリング装置１０６を介してビデオ説明情報３０６Ａのオーディオ表現をレンダリングできる第１の時間間隔（すなわち、t_2～t_3）中に、ビデオ説明情報３０６Ａをディスプレイ装置１０４上に（例えば、テキストフォーマットで）レンダリングすることができる。表２によれば、回路２０２は、第１の時間間隔（すなわち、t_2～t_3）中に、ビデオ説明情報３０６Ａのオーディオ表現（例えば、「Ｊａｃｋが車を運転中、ＧｅｏｒｇｅがＪａｃｋを見ている（

Ｇｅｏｒｇｅ ｉｓ ｌｏｏｋｉｎｇ ａｔ Ｊａｃｋ ｗｈｉｌｅ Ｊａｃｋ ｉｓ ｄｒｉｖｉｎｇ ｔｈｅ ｃａｒ）」）の再生を、撮影シーン３０４のビデオ説明情報３０６Ａのオーディオ表現の実際の再生速度の１．６倍の速度で制御することができる。実際の再生速度は、メディアコンテンツのオーディオをレンダリングできるレート又は速度（すなわち、１倍速）に対応することができる。実際の再生速度は、撮影シーン４０４の取り込み時にオーディオ部分１１６が録音されたレート又は速度とすることができる。ビデオ説明情報３０６Ａのオーディオ表現を再生するための第１の時間間隔（すなわち、図３Ａに示すt_2～t_3）は、テキスト情報３０６に含まれるタイミング情報によって示すことができ、ビデオ説明情報３０６Ａのオーディオ表現を再生できる速度（すなわち、１．６倍）は、テキスト情報３０６に含まれる速度情報によって示すことができる。

【００４９】

限定ではなく一例として、表１によれば、回路２０２は、時間間隔t_A～t_B中に、ビデオ説明情報３０６Ａ（「ビデオ説明１」）のテキスト表現、又はテキスト表現及びオーディオ表現の両方の再生を制御することができる。限定ではなく別の例として、表１によれば、回路２０２は、時間間隔t_A～t_B中に、ビデオ説明情報３０６Ａ（「ビデオ説明１」）のオーディオ表現の再生を、撮影シーン３０４のビデオ説明情報３０６Ａのオーディオ表現の実際の再生速度の０．５倍の速度で制御することができる。従って、開示するメディアレンダリング装置３０２は、ディスプレイ装置１０４及びオーディオレンダリング装置１０６を介して再生できるメディアコンテンツのテキスト情報３０６に（例えば、テキスト形態で）含まれるタイミング情報及び／又は速度情報に基づいて、ビデオ説明情報３０６Ａの（テキスト表現、オーディオ表現、又はテキスト表現及びオーディオ表現の両方での）再生のタイミング及び／又は速度を制御することを可能にすることができる。

【００５０】

ある実施形態では、回路２０２を、メディアコンテンツのレンダリング前又はその最中にユーザ１１２に対してディスプレイ装置１０４上に一連の選択肢を表示するように構成することができる。一連の選択肢のうちの第１の選択肢は、ビデオ説明情報のオーディオ表現（すなわち、ビデオ説明情報をオーディオフォーマットでレンダリングすること）の選択に対応することができる。一連の選択肢のうちの第２の選択肢は、ビデオ説明情報のテキスト表現（すなわち、ビデオ説明情報をテキストフォーマットでレンダリングすること）の選択に対応することができる。同様に、一連の選択肢のうちの第３の選択肢は、ビデオ説明情報のオーディオ表現及びテキスト表現の選択（すなわち、ビデオ説明情報をオーディオ表現及びテキスト表現の両方で同時にレンダリングすること）に対応することができる。いくつかの実施形態では、回路２０２が、ユーザ１２０のユーザプロファイルからビデオ説明情報の再生のためのユーザ選好を決定することができる。回路２０２は、このユーザ選好に基づいてビデオ説明情報の再生（テキストフォーマット、オーディオフォーマット、又はこれらの両方）をさらに制御することができる。

【００５１】

図３Ｂには、ディスプレイ装置１０４及びオーディオレンダリング装置１０６をさらに含むことができるメディアレンダリング装置３０２を示す。メディアレンダリング装置３０２は、ディスプレイ装置１０４及びオーディオレンダリング装置１０６を、メディアコンテンツをレンダリングするように制御することができる。メディアコンテンツは、（図３Ａに示すような）ディスプレイ装置１０４上に表示された撮影シーン３０４を含むことができる（図１の一連の撮影シーン１１４などの）一連の撮影シーンを含むことができる。

【００５２】

ある実施形態では、メディアコンテンツの一連の撮影シーンの各々が、オーディオ部分、画像部分、テキスト情報３０６、及びクローズドキャプション情報３２０を含むことができる。オーディオ部分は、オーディオレンダリング装置１０６を介してレンダリングできるオーディオフォーマットでの、メディアコンテンツの一連の撮影シーンの各々の１又は２以上のせりふを含むことができる。各シーンの画像部分は、ディスプレイ装置１０４上にレンダリングできる１又は２以上の画像フレームを含むことができる。クローズドキ

ャプション情報３２０は、撮影シーン３０４の再生中に（図３Ｂに示すような）ディスプレイ装置１０４上にレンダリングできるテキストフォーマットでの、撮影シーン３０４のオーディオ部分１１６を表すことができる。クローズドキャプション情報３２０は、撮影シーン３０４のオーディオ部分の転写とみなすことができる。いくつかの実施形態では、ビデオ説明情報３０６Ａ（すなわち、シーン説明）をクローズドキャプション情報３２０と共にメディアコンテンツ内に符号化することができる。

【００５３】

ある実施形態では、撮影シーン３０４が、第１のショット及び第２のショットなどの複数のショットを含むことができる。撮影シーン３０４は、複数の生物オブジェクト及び無生物オブジェクトを含むことができる。例えば、図３Ｂでは、生物オブジェクトが、以下に限定するわけではないが、（例えば、「Ｇｅｏｒｇｅ」という名前の）第１の人物３０８及び（例えば、「Ｊａｃｋ」という名前の）第２の人物３１０を含むことができる。図３Ｂに示す無生物オブジェクトは、限定するわけではないが車３１２を含むことができる。

【００５４】

ある実施形態では、図３Ｂに示すように、第１の時間間隔（すなわち、図３Ａに示すt_2～t_3）中に、「Ｊａｃｋが車を運転中、ＧｅｏｒｇｅがＪａｃｋを見ている」というビデオ説明情報３０６Ａをディスプレイ装置１０４上にテキストフォーマットでレンダリングすることができる。別の実施形態では、表２に従って、回路２０２が、第１の時間間隔（t_2～t_3）中に、「Ｊａｃｋが車を運転中、ＧｅｏｒｇｅがＪａｃｋを見ている」というビデオ説明情報３０６Ａのオーディオ表現の再生を、撮影シーン３０４のビデオ説明情報３０６Ａのオーディオ表現の実際の再生速度（すなわち、１．０倍速）の１．６倍の速度で制御することができる。ある実施形態では、図３Ｂに示すように、第１の時間間隔（すなわち、図３Ａに示すt_2～t_3）中に、ビデオ説明情報３０６Ａをクローズドキャプション情報３２０の表示と共にディスプレイ装置１０４上にテキストフォーマットでレンダリングしながら、オーディオレンダリング装置１０６を介してビデオ説明情報３０６Ａのオーディオ表現をレンダリングすることもできる。図３Ｂに示すように、ビデオ説明情報３０６Ａ及びクローズドキャプション情報３２０は、撮影シーン３０４の表示時に（画像フレームなどの）画像部分にオーバーレイ表示できるテキストフォーマットでディスプレイ装置１０４上にレンダリングすることができる。いくつかの実施形態では、ビデオ説明情報３０６Ａのオーディオ表現を再生する代わりに、第１の時間間隔（t_2～t_3）中にビデオ説明情報３０６Ａ及びクローズドキャプション情報３２０を同時にディスプレイ装置１０４上にレンダリングすることができる。

【００５５】

図４は、本開示の実施形態による、シーン説明の再生制御のための別の例示的なシナリオを示す図である。図４の説明は、図１、図２、図３Ａ及び図３Ｂの要素に関連して行う。図４には例示的なシナリオ４００を示す。例示的なシナリオ４００には、メディアレンダリング装置４０２（すなわち、メディアレンダリング装置１０２と同様のもの）を示す。図４には、メディアレンダリング装置４０２に関連するディスプレイ装置１０４及びオーディオレンダリング装置１０６をさらに示す。メディアレンダリング装置４０２は、ディスプレイ装置１０４及びオーディオレンダリング装置１０６を、メディアコンテンツをレンダリングように制御することができる。メディアコンテンツは、ディスプレイ装置１０４上に表示された撮影シーン４０４を含むことができる（図１の一連の撮影シーン１１４などの）一連の撮影シーンを含むことができる。

【００５６】

ある実施形態では、メディアコンテンツの一連の撮影シーンの各々が、オーディオ部分、画像部分、及びビデオ説明情報４０６を含むことができる。いくつかの実施形態では、一連の撮影シーンの各々が、（例えば、図３Ａで説明したようなビデオ説明情報４０６を含むことができるテキスト情報３０６などの）テキスト情報を含むことができる。オーディオ部分は、オーディオレンダリング装置１０６を介してレンダリングできるオーディオフォーマットでの、メディアコンテンツの一連の撮影シーンの各々の１又は２以上のせり

ふを含むことができる。各シーンの画像部分は、ディスプレイ装置１０４上にレンダリングできる１又は２以上の画像フレームを含むことができる。ビデオ説明情報４０６Ａは、一連の撮影シーンのうちの（撮影シーン４０４又は第１の撮影シーン１１４Ａなどの）撮影シーンを説明することができ、撮影シーン４０４はディスプレイ装置１０４上に表示することができる。いくつかの実施形態では、ビデオ説明情報４０６が、一連の撮影シーンの各撮影シーンを説明することができる。

【００５７】

ある実施形態によれば、回路２０２は、メディアレンダリング装置４０２のメモリ２０４から（一連の撮影シーン及びビデオ説明情報４０６を含むことができる）メディアコンテンツを検索するように構成することができる。いくつかの実施形態では、回路２０２を、メディアコンテンツを求める要求をサーバ１０８に送信するように構成することができる。送信された要求に基づいて、サーバ１０８から（一連の撮影シーン及びビデオ説明情報４０６を含むことができる）メディアコンテンツを受け取ることができる。ビデオ説明情報４０６Ａは（オーディオフォーマットではなく）テキストフォーマットで受け取ることができ、これによりサーバ１０８とメディアレンダリング装置４０２との間におけるビデオ説明情報４０６の通信中の帯域幅をさらに節約することができる。テキストフォーマットでのビデオ説明情報４０６は、メモリ２０４又はサーバ１０８の記憶スペースをさらに節約することができる。ある実施形態では、メディアコンテンツの再生前に、メディアコンテンツから一連の撮影シーン及びビデオ説明情報４０６の各々を検索することができる。他のいくつかの実施形態では、回路２０２が、特定の撮影シーン（例えば、現在の撮影シーン）の再生時に、さらなる処理（例えば、次の撮影シーンのビデオ説明情報４０６をオーディオ再生するための速度の計算）のために次の撮影シーンのビデオ説明情報４０６を検索することができる。

【００５８】

回路２０２は、撮影シーン４０４の検索されたビデオ説明情報４０６をビデオ説明情報４０６のオーディオ表現に変換するようにテキスト－スピーチ変換器２０８を制御するようさらに構成することができる。いくつかの実施形態では、撮影シーン４０４に関する情報がクローズドキャプション情報も含む。例えば図３Ｂで説明したように、クローズドキャプション情報は、撮影シーン３０４の表示時に（画像フレームなどの）画像部分にオーバーレイ表示できる、テキストフォーマットでの撮影シーン４０４のオーディオ部分１１６を表すことができる。いくつかの実施形態では、ビデオ説明情報４０６（すなわち、シーン説明）をクローズドキャプション情報と共にメディアコンテンツ内に符号化することができる。

【００５９】

ある実施形態では、撮影シーン４０４が、第１のショット及び第２のショットなどの複数のショットを含むことができる。撮影シーン４０４は、複数の生物オブジェクト及び無生物オブジェクトを含むことができる。例えば、図４では、生物オブジェクトが、以下に限定するわけではないが、（例えば、「Ｇｅｏｒｇｅ」という名前の）第１の人物４０８及び（例えば、「Ｊａｃｋ」という名前の）第２の人物４１０を含むことができる。図４に示す無生物オブジェクトは、限定するわけではないが車３１２を含むことができる。図４にはタイムライン４１４をさらに示す。タイムライン４１４は、撮影シーン４０４を再生できる（例えば、秒、分又は時間単位の）時間間隔を示すことができる。タイムライン４１４に示される合計時間は、撮影シーン４０４に関連する全ての画像フレーム及び／又はオーディオフレームをレンダリングするための再生時間とすることができる。タイムライン４１４は、撮影シーン４０４における第１の人物４０８と第２の人物４１０との間の会話中に発せられるせりふに対応できる一連の第２の時間間隔４１６を含むことができる。

【００６０】

図４に関しては、メディアコンテンツ又は（図１に示すテキスト情報１１８などの）テキスト情報がタイミング情報及び速度情報（すなわち、例えば図３Ａで説明したもの）を含んでいないと仮定することができる。従って、開示するメディアレンダリング装置１０

２は、ビデオ説明情報４０６のオーディオ表現を再生するための速度及び第１の時間間隔を決定することができる。ある実施形態によれば、回路２０２は、（第１のオーディオ部分４１６Ａ及び第２のオーディオ部分４１６Ｂなどの）オーディオ部分１１６を含むことができる撮影シーン４０４の一連の第２の時間間隔４１６を決定するようにさらに構成することができる。一連の第２の時間間隔４１６の各々は、一連の撮影シーンにおける撮影シーン４０４のオーディオ部分１１６を再生するための時間間隔を示すことができる。例えば、図４に示すように、撮影シーン４０４は、撮影シーン４０４のオーディオ部分１１６として第１のオーディオ部分４１６Ａ及び第２のオーディオ部分４１６Ｂを含むことができる。第１のオーディオ部分４１６Ａ及び第２のオーディオ部分４１６Ｂの各々は、撮影シーン４０４に取り込まれた第１の人物４０８及び／又は第２の人物４１０に対応する（図４に示す「Ｇｅｏｒｇｅ：おいＪａｃｋ、どこに向かってるんだ？」及び「Ｊａｃｋ：仕事だよ」などの）せりふを含むことができる。回路２０２は、撮影シーン４０４に含まれる各オーディオフレームのオーディオ分析に基づいて、撮影シーン４０４における一連の第２の時間間隔４１６を決定するように構成することができる。オーディオ分析では、回路２０２が、各オーディオフレーム内のオーディオ音量又はピッチをオーディオ閾値（ｄＢ単位）と比較して、撮影シーン４０４に関連するせりふ又は音楽を含むことができる一連の第２の時間間隔４１６を決定することができる。

【００６１】

ある実施形態では、回路２０２を、撮影シーン４０４のビデオ説明情報４０６のオーディオ表現の第３の時間間隔４１８（すなわち、図４に示すような「t_{00}」～「t_{01}」の時間間隔）を決定するようにさらに構成することができる。第３の時間間隔４１８は、ビデオ説明情報４０６のオーディオ表現をその実際の再生速度でプレイバック又は再生するために必要な期間（例えば、数秒単位）に対応することができる。この時間間隔は、ユーザ１２０がビデオ説明情報４０６を表示する選択肢を選択した場合に（図３Ｂに示すような）ディスプレイ装置１０４上にビデオ説明情報４０６のテキストフォーマットを表示できる期間であることもできる。実際の再生速度は、メディアコンテンツのオーディオをレンダリングできるレート又は速度（すなわち、１倍速）に対応することができる。実際の再生速度は、撮影シーン４０４の取り込み時にオーディオ部分１１６が録音されたレート又は速度とすることができる。ある実施形態では、第３の時間間隔４１８が、ビデオ説明情報４０６のサイズに基づくことができる。例えば、撮影シーン４０４を説明するためにより多くの数の単語がビデオ説明情報４０６に含まれている場合には、ビデオ説明情報４０６のオーディオ表現を実際の再生速度で再生するための第３の時間間隔４１８の期間も長くなることができる。

【００６２】

ある実施形態によれば、回路２０２は、撮影シーン４０４の一連の第４の時間間隔４２０Ａ～４２０Ｃを決定するようにさらに構成することができる。一連の第４の時間間隔４２０Ａ～４２０Ｃの各々は、一連の第２の時間間隔４１６とは異なることができ、撮影シーン４０４のタイムライン４１４内の自然な途切れ（又は空白）に対応できる全ての間隔を含むことができる。図４に示すように、一連の第４の時間間隔４２０Ａ～４２０Ｃは、第１の途切れ４２０Ａ、第２の途切れ４２０Ｂ、及び第３の途切れ４２０Ｃを含むことができる。回路２０２は、撮影シーン４０４に含まれる各オーディオフレームのオーディオ分析に基づいて、撮影シーン４０４内の自然な途切れ又は空白（すなわち、一連の第４の時間間隔４２０Ａ～４２０Ｃに対応する途切れ又は空白）を決定するように構成することができる。オーディオ分析では、回路２０２が、各オーディオフレーム内のオーディオ音量又はピッチをオーディオ閾値（ｄＢ単位）と比較することができる。オーディオフレーム内のオーディオ音量又はピッチが（例えば、ｄＢ単位の）オーディオ閾値よりも小さい場合には、対応するオーディオフレームを撮影シーン４０４内の自然な途切れ又は空白として決定することができる。回路２０２は、撮影シーン４０４に含まれる第１の途切れ４２０Ａ、第２の途切れ４２０Ｂ又は第３の途切れ４２０Ｃなどの決定された途切れ又は空白を再生するための一連の第４の時間間隔４２０Ａ～４２０Ｃ又は期間を決定するように

さらに構成することができる。

【００６３】

一例として、図４に示すように、撮影シーン４０４に関連するタイムライン４１４には、時点t_0～t_1に、第１の途切れ４２０Ａに対応できる自然な途切れが存在することができる。時点t_1～t_2には、第１の人物４０８が、せりふ又はオーディオ部分１１６として「おいＪａｃｋ、どこに向かってるんだ？」と発話することができる。さらに、時点t_2～t_3には、第２の途切れ４２０Ｂに対応できる別の自然な途切れが存在することができる。時点t_3～t_4には、第２の人物４１０が、例えば「仕事だよ」というせりふで第１の人物４０８に返答することができる。時点t_4～t_5には、第３の途切れ４２０Ｃに対応できる別の自然な途切れが存在することができる。従って、図４に示すように、一連の第２の時間間隔４１６は、時点t_1からt_2に及ぶことができる第１のオーディオ部分４１６Ａ、及び時点t_3からt_4に及ぶことができる第２のオーディオ部分４１６Ｂを含むことができる。一連の第４の時間間隔４２０Ａ～４２０Ｃは、時点t_0からt_1に及ぶことができる第１の途切れ４２０Ａ、時点t_2からt_3に及ぶことができる第２の途切れ４２０Ｂ、及び時点t_4からt_5に及ぶことができる第３の途切れ４２０Ｃを含むことができる。

【００６４】

回路２０２は、撮影シーン４０４の一連の第４の時間間隔４２０Ａ～４２０Ｃから第１の時間間隔４２２を選択するようにさらに構成することができる。第１の時間間隔４２２は、時間間隔閾値の期間よりも長い期間を有することができる時間間隔であることができ、ビデオ説明情報４０６（すなわち、シーン説明）のオーディオ再生のための潜在的空白とみなすことができる。時間間隔閾値は、第１の人物４０８又は第２の人物４１０が特定のせりふを発話している間に発生し得る短い途切れ又は空白をフィルタ除去するために利用される（例えば、ミリ秒又は数秒単位の）所定の時間値とすることができる。例えば、時間間隔閾値は、第１の人物４０８又は第２の人物４１０が複数のせりふ間に息を吸う／吐くために要する時間を示すことができる。

【００６５】

ある実施形態では、回路２０２が、一連の第４の時間間隔４２０Ａ～４２０Ｃの各々と時間間隔閾値との比較に基づいて第１の時間間隔４２２を選択することができる。時間間隔閾値は、ビデオ説明情報４０６のオーディオ再生が不可能と考えられる間隔の値に対応することができる。換言すれば、時間間隔閾値は、それ未満ではビデオ説明情報４０６のオーディオ再生がメディアコンテンツのレンダリング対象であるユーザ１２０に対して十分に詳細なシーン説明を提供できないと考えられるタイミング値に対応することができる。

【００６６】

例えば、第１の途切れ４２０Ａの期間が０．７５秒であり、第２の途切れ４２０Ｂの期間が１秒であり、第３の途切れ４２０Ｃの期間が０．５秒であり、時間間隔閾値が１秒である場合、回路２０２は、一連の第４の時間間隔４２０Ａ～４２０Ｃ内の各途切れの期間と時間間隔閾値とを比較し、時間間隔閾値以上の期間を有する第２の途切れ４２０Ｂを第１の時間間隔４２２として選択することができる。いくつかの実施形態では、期間が長くなるとビデオ説明情報４０６（すなわち、シーン説明）の再生速度が実際の再生速度と同じになり、従ってビデオ説明情報４０６のオーディオ再生の品質を維持することができるので、回路２０２は、（第１の途切れ４２０Ａ、第２の途切れ４２０Ｂ、又は第３の途切れ４２０Ｃのうちの）最も長い期間を有する途切れを第１の時間間隔４２２として選択することができる。

【００６７】

ある実施形態によれば、回路２０２は、ビデオ説明情報４０６のオーディオ表現を再生する再生速度を決定するようにさらに構成することができる。再生速度は、ビデオ説明情報４０６のオーディオ表現の再生速度に対応することができる。いくつかの実施形態では、回路２０２が乗算係数（ｍｕｌｔｉｐｌｉｃａｔｉｏｎ　ｆａｃｔｏｒ）を計算し、計算された乗算係数及びビデオ説明情報４０６のオーディオ表現の実際の再生速度に基づいて再生速度を決定することができる。乗算係数は、決定された第３の時間間隔４１８及び

選択された第１の時間間隔４２２に基づいて計算することができる。

【００６８】

ある例では、撮影シーン４０４内の第１の途切れ４２０Ａ（時点t_0～t_1）の期間が２秒であり、第２の途切れ４２０Ｂ（時点t_2～t_3）の期間が３秒であり、第３の途切れ４２０Ｃ（時点t_4～t_5）の期間が２秒である。第３の時間間隔４１８の期間が５秒である場合、このような期間は、一連の第４の時間間隔４２０Ａ～４２０Ｃ（すなわち、第１の途切れ４２０Ａ、第２の途切れ４２０Ｂ、及び第３の途切れ４２０Ｃ）の各々又は選択された第１の時間間隔４２２に対応する時間間隔中にビデオ説明情報４０６を実際の再生速度で聞き取れるように再生するには不十分と考えられる。回路２０２は、以下の方程式（１）を使用して乗算係数を決定するように構成することができる。

$$\text{乗算係数} = \frac{\text{第３の時間間隔}}{\text{第１の時間間隔}} \qquad (1)$$

【００６９】

回路２０２は、計算された乗算係数及び実際の再生速度に基づいて、以下の方程式（２）を使用することによって、ビデオ説明情報４０６のオーディオ表現を再生する再生速度を決定するようにさらに構成することができる。

再生速度＝乗算係数*実際の再生速度（２）

【００７０】

上述した例を参照すると、回路２０２は、方程式（１）を使用することにより、乗算係数を１．６６（すなわち、５秒である第３の時間間隔４１８と、３秒である第２の途切れ４２０Ｂとして選択された第１の時間間隔４２２との比率）であると決定するように構成することができる。乗算係数が１．０よりも大きい（すなわち、第３の時間間隔４１８が第１の時間間隔４２２よりも大きい）場合、回路２０２は、ビデオ説明情報４０６のオーディオ表現の実際の再生速度を乗算係数によって増加させるように構成することができる。例えば、乗算係数が１．６６である場合、回路２０２は、撮影シーン４０４のビデオ説明情報４０６のオーディオ表現の実際の再生速度の１．６６倍を再生速度として決定することができる。その他の事例では、乗算係数が１．０未満である場合（すなわち、第３の時間間隔４１８が第１の時間間隔４２２よりも小さい場合）、回路２０２は、ビデオ説明情報４０６のオーディオ表現の実際の再生速度を乗算係数によって減少させるように構成することができる。例えば、乗算係数が０．８である場合、回路２０２は、撮影シーン４０４のビデオ説明情報４０６のオーディオ表現の実際の再生速度の０．８倍を再生速度として決定することができる。いくつかの実施形態では、乗算係数が１．０未満である場合、回路２０２は実際の再生速度を変更せず、ビデオ説明情報４０６のオーディオ表現の再生速度は実際の再生速度と同じままであることができる（例えば、乗算係数が０．９５である場合には実質的に１．０に近いと考えることができる）。他のいくつかの実施形態では、乗算係数が１．０に等しい場合（すなわち、第３の時間間隔４１８が第１の時間間隔４２２に等しい場合）、回路２０２は、ビデオ説明情報４０６のオーディオ表現の実際の再生速度を再生速度として決定するように構成することができる。

【００７１】

回路２０２は、決定された再生速度に基づいて、ビデオ説明情報４０６のオーディオ表現の再生を第１の時間間隔４２２において制御するようにさらに構成することができる。第１の時間間隔４２２（すなわち、途切れのうちの１つ）は、一連の第２の時間間隔４１６（すなわち、撮影シーン４０４のオーディオ部分を含む第２の時間間隔）とは異なることができる。いくつかの実施形態では、第１の時間間隔４２２を、撮影シーン４０４の第１のせりふ（例えば、第１のオーディオ部分４１６Ａ）と第２のせりふ（例えば、第２のオーディオ部分４１６Ｂ）との間とすることができる。例えば、図４に示すように、ビデオ説明情報４０６のオーディオ表現（すなわち、「Ｊａｃｋが車を運転中、Ｇｅｏｒｇｅ

がＪａｃｋを見ている」というシーン説明）は、第１の時間間隔４２２において、決定された再生速度で（例えば、第３の時間間隔４１８が５秒であり、第１の時間間隔４２２が３秒である場合には１．６６倍で）再生することができる。従って、回路２０２は、ビデオ説明情報４０６の一部（例えば、特定の文字、テキスト又は単語）を短縮又は削除することなく、オーディオせりふの空白（すなわち、第１の時間間隔４２２）間のビデオ説明情報４０６（すなわち、シーン説明）のオーディオ再生速度を増加させることができる。この速度の増加により、ビデオ説明情報４０６の第３の時間間隔４１８よりも短い期間である第１の時間間隔４２２内にビデオ説明情報４０６のオーディオ表現を効果的に組み込み又は収めることができる。従って、たとえ決定された空白（すなわち、撮影シーン４０４内の特定の空白の第１の時間間隔４２２）が第３の時間間隔４１８（すなわち、シーン／ビデオ説明を聞き取れるように再生するのに必要な時間）より短い場合でも、ビデオ説明情報４０６の再生品質が維持される。

【００７２】

いくつかの実施形態では、第１のせりふを撮影シーン４０４の第１のショットの最後の単語とすることができ、第２のせりふを撮影シーン４０４の第２のショットの最初の単語とすることができる。第１のショット及び第２のショットは、撮影シーン４０４の連続するショットとすることができる。他のいくつかの実施形態では、第１の時間間隔４２２を、撮影シーン４０４の開始と撮影シーン４０４の（第１のオーディオ部分４１６Ａなどの）第１のせりふとの間とすることができる。このような場合、第１の時間間隔４２２は、図４に示すような第１の途切れ４２０Ａに対応することができる。

【００７３】

なお、図４に示す撮影シーン４０４、及び複数の生物オブジェクト又は無生物オブジェクトは、一例として提示するものにすぎない。本開示は、他のタイプの撮影シーン（例えば、以下に限定するわけではないが、アクションシーン、恋愛シーン、ドラマシーン、ダンスシーン又は音楽シーン）及び複数の生物オブジェクト又は無生物オブジェクトにも適用可能である。他のタイプの撮影シーン４０４及び複数の生物オブジェクト又は無生物オブジェクト、或いはこれらの例の説明については、簡潔にするために本開示からは省略する。

【００７４】

ある実施形態では、回路２０２を、一連の撮影シーン１１４の各々について、対応する撮影シーンのオーディオ部分を再生するための時間間隔をそれぞれが示すことができる一連の第２の時間間隔を決定するように構成することができる。回路２０２は、一連の撮影シーン１１４のうちの対応する撮影シーンのビデオ説明情報のオーディオ表現の第３の時間間隔を決定するようにさらに構成することができる。撮影シーン４０４に関して上述したように、回路２０２は、各シーンの決定された一連の第２の時間間隔及び決定された第３の時間間隔に基づいて、ビデオ説明情報４０６のオーディオ表現を再生する速度を決定するようにさらに構成することができる。回路２０２は、決定された速度に基づいて、一連の撮影シーン１１４の各撮影シーンのビデオ説明情報のオーディオ表現の再生を第１の時間間隔（すなわち、一連の第２の時間間隔とは異なる時間間隔）において制御するようにさらに構成することができる。従って、開示するメディアレンダリング装置４０２は、対応する撮影シーン又は以前の撮影シーン（すなわち、対応する撮影シーンの直前のシーン）の再生中に、メディアコンテンツ内の各撮影シーンを処理し、対応する撮影シーンの第１の時間間隔４２２を選択し、撮影シーンに関連するビデオ説明情報４０６の再生速度を決定することができる。さらに、メディアレンダリング装置４０２は、一連の撮影シーン１１４内の各撮影シーンの決定された再生速度に基づいて、対応するビデオ説明情報のオーディオ表現（すなわち、シーン説明）の再生を動的に制御することができる。従って、開示するメディアレンダリング装置４０２は、例えば視覚障害者又は認知障害者などのユーザ１２０のコンテンツ体験を強化することができる。

【００７５】

ある実施形態では、回路２０２を、ユーザ１２０からＩ／Ｏ装置２０６を介して第１の

ユーザ入力を受け取るようにさらに構成することができる。第１のユーザ入力はテキストフォーマットであることができ、ビデオ説明情報４０６、又は一連の撮影シーン１１４のうちの１つの撮影シーンのシーン説明に対応することができる。回路２０２は、メディアコンテンツの再生中又はメディアコンテンツの再生開始前に第１のユーザ入力を受け取ることができる。第１のユーザ入力は、一連の撮影シーン１１４のうちの撮影シーンのうちの１つの撮影シーンのビデオ説明情報４０６に含めることができるテキスト単語又は表現とすることができる。例えば、図４に示すビデオ説明情報４０６は、「Ｊａｃｋが車を運転中、ＧｅｏｒｇｅがＪａｃｋを見ている」であることができる。受け取られた第１のユーザ入力は、ビデオ説明情報４０６の一部であることができる単語又は表現（例えば、「ＧｅｏｒｇｅがＪａｃｋを見ている」）を含むことができる。

【００７６】

回路２０２は、一連の撮影シーン１１４の各々に関連する記憶されたビデオ説明情報４０６内で、受け取られた第１のユーザ入力を検索するようにさらに構成することができる。いくつかの実施形態では、第１のユーザ入力で受け取られたテキスト説明が、一連の撮影シーン１１４のうちの１つの撮影シーンのビデオ説明情報４０６と全く同じものであることができる。他の実施形態では、第１のユーザ入力が、ビデオ説明情報４０６の一部であることができる。回路２０２は、検索に基づいて、メディアコンテンツを再生するための再生タイミング情報を決定するようにさらに構成することができる。回路２０２は、検索に基づいて再生タイミング情報を決定するために、撮影シーン（例えば、撮影シーン４０４）、及び第１のユーザ入力を含む対応するビデオ説明情報４０６を決定することができる。このような場合、再生タイミング情報は、決定された撮影シーンの再生タイミングであることができる。他のいくつかの実施形態では、第１のユーザ入力で受け取られたテキスト説明が、一連の撮影シーン１１４の各々に関連するビデオ説明情報４０６と全く同じではないことがある。このようなシナリオでは、回路２０２を、第１のユーザ入力において受け取られたテキスト説明と、一連の撮影シーン１１４の各々に関連するビデオ説明情報４０６との間の類似性スコアを決定するように構成することができる。類似性スコアは、テキスト説明と対応する撮影シーンのビデオ説明情報４０６の部分との一致に基づいて決定することができる。いくつかの実施形態では、類似性スコアを、メディアコンテンツの一連の撮影シーン１１４の各々に関連する人気度スコアに基づいて計算することができる。回路２０２は、サーバ１０８から各撮影シーンの人気スコアを検索することができる。ある実施形態では、サーバ１０８又はメモリ２０４から検索されたメディアコンテンツに各撮影シーンの人気スコアを含めることができる。例えば、第１のユーザ入力（すなわち、説明）が「ＧｅｏｒｇｅがＪａｃｋを見ている」という単語であり、この単語が、撮影シーン４０４を含む複数の撮影シーンに関連するビデオ説明情報４０６内に存在し得るものとする。このような場合、回路２０２は、複数の撮影シーンの中の各撮影シーンの人気スコアを抽出し、どのシーンが人々の間で人気が高く、ユーザ１２０が人気の高い撮影シーンのビデオ説明情報４０６の説明を検索したいと思っている確率が高い（例えば、撮影シーン４０４）のはどのシーンであるかを識別することができる。回路２０２は、受け取られた説明（又は第１のユーザ入力）の類似度スコアが高い識別された撮影シーン（例えば、撮影シーン４０４）の再生タイミング情報を決定するようにさらに構成することができる。回路２０２は、決定された再生タイミング情報（t_0）に基づいて、識別された撮影シーンからのメディアコンテンツの再生を制御するようにさらに構成することができる。従って、開示するメディアレンダリング装置４０２は、ユーザ１２０がメディアコンテンツの一連の撮影シーン１１４の各々の記憶されたビデオ説明情報４０６（すなわち、シーン説明）内の単語又はテキストを検索し、従って検索に基づいて識別できる識別された撮影シーンの再生タイミングを制御（すなわち、早送り又は巻き戻し）することを可能にすることができる。従って、メディアレンダリング装置４０２は、メディアコンテンツ内の１又は２以上のシーンに対応する説明をユーザ１２０が検索できるようにする検索エンジン機能を提供することができる。

【００７７】

ある実施形態では、メディアレンダリング装置４０２を、一定期間（例えば、最後の１日又は１週間）内に第１のユーザ入力で受け取られた以前の検索説明に基づいて、新たなメディアコンテンツの個人化された推奨を提供するようにさらに構成することができる。一例として、ユーザ１２０が特定の期間内に「アクション」という単語を検索した頻度が高い場合、回路２０２は、「アクション」ジャンルに関連し得る他の又は新たなメディアコンテンツの推奨を提供することができる。従って、開示するメディアレンダリング装置４０２は、ユーザ１２０が頻繁に検索していると考えられるシーン又はビデオ説明に関連するメディアコンテンツを推奨することができる。

【００７８】

ある実施形態では、メディアレンダリング装置４０２を、メディアレンダリング装置４０２に関連する第１の規定の速度設定をメモリ２０４に記憶するように構成することができる。第１の規定の速度設定は、ビデオ説明情報４０６（すなわち、シーン説明）のオーディオ表現の最大再生速度を示すことができる。第１の規定の速度設定によって示される最大速度は、メディアコンテンツのレンダリング対象であるユーザ１２０がビデオ説明情報４０６のオーディオ表現を正しく理解できる速度とすることができる。例えば、最大速度は、実際の再生速度の２倍とすることができる。いくつかの実施形態では、第１の規定の速度設定が、再生速度を決定できる基になる乗算係数の最大値（例えば、２．０）を示すことができる。

【００７９】

別の実施形態では、メディアレンダリング装置４０２を、メディアレンダリング装置４０２に関連する第２の規定の速度設定をメモリ２０４に記憶するように構成することができる。第２の規定の速度設定は、ビデオ説明情報４０６（すなわち、シーン説明）のオーディオ表現の最小再生速度を示すことができる。第２の規定の速度設定によって示される最小速度は、メディアコンテンツのレンダリング対象であるユーザ１２０がビデオ説明情報４０６のオーディオ表現を正しく理解できる速度とすることができる。例えば、最小速度は、実際の再生速度の０．５倍とすることができる。いくつかの実施形態では、第２の規定の速度設定が、再生速度を決定できる基になる乗算係数の最小値（例えば、０．５）を示すことができる。

【００８０】

ある実施形態によれば、回路２０２を、ビデオ説明情報４０６のオーディオ表現の決定された再生速度、及び第１／第２の規定の速度設定に基づいて、撮影シーン４０４の画像部分又はオーディオ部分の一方の再生を制御するようにさらに構成することができる。撮影シーン４０４の画像部分又はオーディオ部分（すなわち、せりふ）の一方の再生制御は、自然な途切れ（すなわち、第１の時間間隔４２２）が、決定された再生速度及び第１又は第２の規定の速度設定に基づいてビデオ説明情報４０６のオーディオ表現を収めることができるほど十分に長い場合の、撮影シーンの画像部分及び／又はオーディオ部分のレンダリングの時間遅延又は一時停止に対応することができる。

【００８１】

一例として、第１の規定の速度設定（すなわち、最大速度）がビデオ説明情報４０６のオーディオ表現の実際の再生速度の２倍であり、第３の時間間隔４１８が７秒であり、第１の時間間隔４２２の期間が３秒である場合、方程式（１）によれば、決定された再生速度は２．３３倍となる。決定された再生速度が最大速度（すなわち、２倍）よりも高いので、回路２０２は、一連の第４の時間間隔４２０Ａ～４２０Ｃから選択された第１の時間間隔４２２を廃棄することができる。このような場合、回路２０２は、ビデオ説明情報４０６のオーディオ表現をレンダリングするために撮影シーン４０４の画像部分又はオーディオ部分（すなわち、図４の第２のオーディオ部分４１６Ｂなどのせりふ）を一時停止することができる。別の事例では、回路２０２が、レンダリングされているメディアコンテンツの品質を維持するために、ビデオ説明情報４０６を（２倍のような）最大速度で聞き取れるようにレンダリングし、撮影シーンの画像部分又はオーディオ部分を（第３の時間間隔４１８が７秒であり、第１の時間間隔４２２の期間が３秒である場合の残りの１秒な

どの）残りの時間にわたって一時停止することができる。
【００８２】
ある実施形態では、回路２０２を、Ｉ／Ｏ装置２０６を介してユーザ１２０から第２のユーザ入力を受け取るようにさらに構成することができる。第２のユーザ入力は、メディアコンテンツをレンダリングできる対象であるユーザ１２０のプロファイル情報を示すことができる。プロファイル情報は、ビデオ説明情報４０６を聞き取れるようにレンダリングするためのユーザ１２０の過去の速度選好を含むことができる。いくつかの実施形態では、プロファイル情報が、ユーザ１２０に関連し得る一意の識別番号（例えば、以下に限定するわけではないが、社会保障番号（ＳＳＮ）、電話番号、又は保険証券番号）を示すことができる。回路２０２は、受け取られた一意の識別番号に基づいて、サーバ１０８又はメモリ２０４からユーザ１２０の年齢を検索するようにさらに構成することができる。いくつかの実施形態では、回路２０２を、ユーザ１２０に関連する一意の識別番号に基づいてユーザ１２０の健康状態を決定するようにさらに構成することができる。健康状態は、ビデオ説明情報４０６のオーディオ表現又は撮影シーンのオーディオ部分（すなわち、せりふ）を特定の再生速度で理解するためのユーザ１２０の聞き取り能力の欠如を示すことができる。回路２０２は、受け取られた第２のユーザ入力に基づいて、ビデオ説明情報４０６のオーディオ表現を再生する再生速度を決定するようにさらに構成することができる。
【００８３】
一例として、ユーザ１２０の年齢が６５歳（すなわち、老齢）として決定された場合、回路２０２は、ビデオ説明情報４０６のオーディオ表現の実際の再生速度の１．５倍を再生速度として決定することができる。いくつかの実施形態では、回路２０２が、決定された年齢に基づいて（例えば、１．５倍を最大速度とする）第１の速度設定を定めることができる。別の例として、ユーザ１２０の健康状態によってユーザ１２０が過去の所定の期間内（例えば、過去６ヶ月以内）に耳の手術を受けたことが示される場合、回路２０２は、ビデオ説明情報４０６のオーディオ表現の実際の再生速度の１．２倍を第１の速度設定として定め、又は再生速度として決定することができる。従って、開示するメディアレンダリング装置４０２は、ユーザ１２０の（年齢又は健康状態などの）プロファイル情報に基づいて、視覚障害又は聴覚障害問題の一方又は両方を有する可能性がある異なるユーザにとってオーディオシーン説明の再生品質が維持されるように、シーン／ビデオ説明を再生するための再生速度又は速度設定（例えば、最大又は最小）を制御することができる。
【００８４】
ある実施形態では、メディアレンダリング装置４０２のメモリ２０４に（図２に示す）訓練済み機械学習（ＭＬ）モデル２１２を記憶することができる。訓練済みＭＬモデル２１２は、撮影シーン４０４のコンテキスト情報（すなわち、コンテキストを示す情報）に基づいて、ビデオ説明情報４０６のオーディオ表現を再生するためのオーディオ特性を決定又は出力することができる。コンテキスト情報は、訓練済み機械学習（ＭＬ）モデル２１２への入力であることができる。機械学習（ＭＬ）モデル２１２は、入力（すなわち、コンテキスト情報）と出力（すなわち、オーディオ特性）との間の関係を識別するように訓練することができる。ＭＬモデル２１２は、例えば重みの数、コスト関数、入力サイズ及び層の数などのハイパーパラメータによって定めることができる。ＭＬモデル２１２のハイパーパラメータは、ＭＬモデル２１２のコスト関数の大域的最小点に近づくように調整することができ、重みもそのように更新することができる。ＭＬモデル２１２は、ＭＬモデル２１２の訓練データセット内の特徴に基づく数エポックの訓練後に一連の入力（すなわち、コンテキスト情報）に対して予測結果（例えば、オーディオ特性）を出力するように訓練することができる。
【００８５】
ＭＬモデル２１２は、例えばソフトウェアプログラム、ソフトウェアプログラムのコード、ライブラリ、アプリケーション、スクリプト、或いは回路２０２などの処理装置によって実行されるその他のロジック又は命令などの電子データを含むことができる。ＭＬモ

デル２１２は、メディアレンダリング装置４０２などのコンピュータ装置がコンテキスト情報に基づいてオーディオ特性を決定するための１又は２以上の動作を実行することを可能にするように構成されたコード及びルーチンを含むことができる。これに加えて又はこれに代えて、ＭＬモデル２１２は、プロセッサ、（例えば、１又は２以上の動作を実行し又は実行を制御する）マイクロプロセッサ、フィールドプログラマブルゲートアレイ（ＦＰＧＡ）、又は特定用途向け集積回路（ＡＳＩＣ）を含むハードウェアを使用して実装することもできる。或いは、いくつかの実施形態では、ハードウェアとソフトウェアとの組み合わせを使用してＭＬモデル２１２を実装することもできる。

【００８６】

ある実施形態によれば、回路２０２は、撮影シーン４０４のコンテキスト情報を決定するように構成することができる。コンテキスト情報の例としては、以下に限定するわけではないが、アクション、格闘、冒険、アニメーション、コメディ、ダンス、ミュージカル、犯罪、叙事詩、エロティカ、ファンタジー、ホラー、ミステリー、哲学、政治、宗教、ロマンス、ＳＦ、スリラー、都市、戦争、伝記、又は悲劇を挙げることができる。コンテキスト情報は、撮影シーン４０４の少なくとも１つの視覚的特性の分析に基づいて決定することができる。撮影シーン４０４の視覚的特性としては、以下に限定するわけではないが、少なくとも１つのフレーム内で認識される物体（例えば、図４の車４１２）、少なくとも１つのフレーム内で認識される（図４の第１の人物４０８又は第２の人物４１０などの）人物、少なくとも１つのフレーム内の少なくとも１つのオブジェクトの（幸福状態、悲しみ状態、怒り状態、混乱状態、ストレス状態、又は興奮状態などの）感情状態、少なくとも１つのフレームの背景情報、少なくとも１つのフレーム内の周囲照明条件、少なくとも１つのフレーム内の動き情報（すなわち、静止又は移動）、少なくとも１つのフレーム内の少なくとも１つのオブジェクトに関連する（ダンスジェスチャ又はアクションジェスチャなどの）ジェスチャ、又は少なくとも１つのフレームに関連するジャンル情報を挙げることができる。いくつかの実施形態では、回路２０２を、（撮影シーン４０４などの）撮影シーンの視覚的特徴及びコンテキスト情報を決定するために、当業で周知の様々な画像処理法、シーンマイニング法、又はシーン理解法を実装するように構成することができる。

【００８７】

回路２０２は、撮影シーン４０４の決定されたコンテキストに対する訓練済みＭＬモデル２１２の適用に基づいて、ビデオ説明情報４０６のオーディオ表現を再生するためのオーディオ特性を決定するようにさらに構成することができる。オーディオ特性は、以下に限定するわけではないが、ラウドネスパラメータ、ピッチパラメータ、トーンパラメータ、発話速度パラメータ、声質パラメータ、音声学的パラメータ、イントネーションパラメータ、倍音の強度、音声変調パラメータ、発音パラメータ、韻律パラメータ、音色パラメータ、或いは１又は２以上の音響心理的パラメータを含むことができる。オーディオ特性は、撮影シーン４０４の決定されたコンテキスト情報に対する訓練済みＭＬモデル２１２の適用に基づいて決定することができる。

【００８８】

回路２０２は、決定された速度及び決定されたオーディオ特性に基づいて、ビデオ説明情報４０６のオーディオ表現の再生を第１の時間間隔４２２において制御するようにさらに構成することができる。一例として、撮影シーン４０４のコンテキスト情報が格闘シーンとして決定された場合、回路２０２は、メディアコンテンツ及びビデオ説明情報４０６（すなわち、シーン説明）をレンダリングできる対象であるユーザ１２０にリアルなユーザ体験を提供するために、ビデオ説明情報４０６の（音量などの）ラウドネスパラメータ、及び倍音パラメータ（すなわち、オーディオ特性）の強度を高めるように構成することができる。このような場合、回路２０２は、コンテキスト情報が格闘シーンとして決定されたことに基づいて、決定されたコンテキスト情報に対する訓練済みＭＬモデル２１２の適用に基づいて（音量などの）ラウドネスパラメータをオーディオ特性として決定することができる。

【００８９】
別の実施形態では、回路２０２を、ビデオ説明情報４０６のオーディオ表現、並びに撮影シーン４０４又は一連の撮影シーンの各撮影シーンの（第１のオーディオ部分４１６Ａ及び第２のオーディオ部分４１６Ｂなどの）オーディオ部分を聞き取れるように再生するようにオーディオレンダリング装置１０６を制御するようさらに構成することができる。オーディオレンダリング装置１０６は、（図２に示すような）メディアレンダリング装置４０２に関連することができ、又はメディアレンダリング装置４０２内に統合することができる。
【００９０】
ある実施形態では、撮影シーン４０４のビデオ説明情報４０６が、撮影シーン４０４内に存在する生物オブジェクト及び／又は無生物オブジェクトに関する認知情報を含むことができる。生物オブジェクトは（人間、動物又は鳥などの）生物を含むことができる。無生物オブジェクトは無生物を含むことができる。オブジェクト（生物又は無生物）に関する認知情報は、撮影シーン４０４のコンテキストに関連することも又はしないこともあるオブジェクトの徹底的な詳細を提供することができる。認知情報は、撮影シーン４０４内に存在するオブジェクトに関する一般的知識又は情報をユーザ１２０に提供することができる。いくつかの実施形態では、認知情報が、撮影シーン内に存在するオブジェクトに関連する画像又はアイコンに対応することができ、或いはオブジェクトに関連するオーディオトーンに対応することができる。ある実施形態では、回路２０２を、ディスプレイ装置１０４又はオーディオレンダリング装置１０６のいずれかによる認知情報の再生を制御するようにさらに構成することができる。
【００９１】
図５は、本開示の実施形態による、シーン説明の再生制御のための例示的な動作を示す第１のフローチャートである。図５の説明は、図１、図２、図３Ａ、図３Ｂ及び図４の要素に関連して行う。図５にはフローチャート５００を示す。５０２～５０８の動作は、例えばメディアレンダリング装置１０２又は回路２０２などのいずれかのコンピュータ装置上で実施することができる。動作は５０２から開始して５０４に進むことができる。
【００９２】
５０４において、メディアコンテンツを検索することができる。メディアコンテンツは、一連の撮影シーン１１４及びテキスト情報１１８を含むことができる。テキスト情報１１８は、ビデオ説明情報１１８Ａ及びタイミング情報１１８Ｂを含むことができる。ビデオ説明情報１１８Ａは、一連の撮影シーン１１４の撮影シーンを説明することができる。１又は２以上の実施形態では、回路２０２を、一連の撮影シーン１１４及びテキスト情報１１８を含むことができるメディアコンテンツ１１２を検索するように構成することができる。テキスト情報１１８は、ビデオ説明情報１１８Ａ及びタイミング情報１１８Ｂをさらに含むことができる。ビデオ説明情報１１８Ａは、一連の撮影シーン１１４の撮影シーンを説明することができる。
【００９３】
５０６において、撮影シーンのテキスト情報１１８から、ビデオ説明情報１１８Ａを再生するためのタイミング情報１１８Ｂを抽出することができる。１又は２以上の実施形態では、回路２０２を、撮影シーンのテキスト情報１１８からタイミング情報１１８Ｂ（すなわち、ビデオ説明情報１１８Ａを再生するためのタイミング情報）を抽出するように構成することができる。
【００９４】
５０８において、ビデオ説明情報１１８Ａ（テキスト表現、オーディオ表現、又はテキスト表現及びオーディオ表現の両方）の再生を制御することができる。ビデオ説明情報１１８Ａは、撮影シーンの抽出されたタイミング情報１１８Ｂによって示される第１の時間間隔において再生することができる。１又は２以上の実施形態では、回路２０２を、撮影シーンの抽出されたタイミング情報によって示される第１の時間間隔においてビデオ説明情報１１８Ａの再生（テキスト表現、オーディオ表現、又はテキスト表現及びオーディオ

表現の両方）を制御するように構成することができる。制御は終了に進むことができる。
【００９５】
図６は、本開示の実施形態による、シーン説明の再生制御のための例示的な動作を示す第２のフローチャートである。図６の説明は、図１、図２、図３Ａ、図３Ｂ、図４及び図５の要素に関連して行う。図６にはフローチャート６００を示す。６０２～６１０の動作は、例えばメディアレンダリング装置１０２又は回路２０２などのいずれかのコンピュータ装置上で実施することができる。動作は６０２から開始して６０４に進むことができる。
【００９６】
６０４において、第１の撮影シーン１１４Ａの一連の第２の時間間隔を決定することができる。一連の第２の時間間隔の各々は、一連の撮影シーン１１４における撮影シーンのオーディオ部分１１６を再生するための時間間隔を示すことができる。１又は２以上の実施形態では、回路２０２を、一連の撮影シーンにおける撮影シーンのオーディオ部分１１６を再生するための時間間隔をそれぞれが示すことができる、撮影シーンの一連の第２の時間間隔を決定するように構成することができる。一連の第２の時間間隔の決定の詳細については、例えば図４で説明している。
【００９７】
６０６において、撮影シーンのビデオ説明情報１１８Ａのオーディオ表現の第３の時間間隔を決定することができる。１又は２以上の実施形態では、回路２０２を、撮影シーンのビデオ説明情報１１８Ａのオーディオ表現の第３の時間間隔を決定するように構成することができる。第３の時間間隔の決定の詳細については、例えば図４で説明している。
【００９８】
６０８において、決定された一連の第２の時間間隔及び決定された第３の時間間隔に基づいて、ビデオ説明情報１１８Ａのオーディオ表現を再生する速度を決定することができる。１又は２以上の実施形態では、回路２０２を、決定された一連の第２の時間間隔及び決定された第３の時間間隔に基づいて、ビデオ説明情報１１８Ａ（すなわち、シーン説明）のオーディオ表現を再生する速度を決定するように構成することができる。ビデオ説明情報の再生速度の決定に関する詳細については、例えば図４で説明している。
【００９９】
６１０において、決定された速度に基づいてビデオ説明情報１１８Ａのオーディオ表現の再生を制御することができる。ビデオ説明情報１１８Ａのオーディオ表現は、一連の第２の時間間隔とは異なることができる第１の時間間隔において再生することができる。１又は２以上の実施形態では、回路２０２を、決定された速度に基づいてビデオ説明情報１１８Ａのオーディオ表現の再生を第１の時間間隔において制御するように構成することができる。ビデオ説明情報１１８Ａのオーディオ表現の再生を制御する詳細については、例えば図４で説明している。制御は終了に進むことができる。
【０１００】
本開示の様々な実施形態は、メディアレンダリング装置４０２などの機械及び／又はコンピュータが実行できる命令を記憶した非一時的コンピュータ可読媒体及び／又は記憶媒体を提供することができる。これらの命令は、一連の撮影シーンを含むことができるメディアコンテンツの検索を含むことができる動作を機械及び／又はコンピュータに実行させることができる。メディアコンテンツは、ビデオ説明情報及びタイミング情報を含むテキスト情報を含むことができる。ビデオ説明情報は、一連の撮影シーン内の撮影シーンを説明することができる。動作は、撮影シーンのテキスト情報から、ビデオ説明情報を再生するためのタイミング情報を抽出することをさらに含むことができる。動作は、ビデオ説明情報の再生を、抽出された撮影シーンのタイミング情報によって示される第１の時間間隔においてテキスト表現又はテキスト表現及びオーディオ表現のいずれかで制御することをさらに含むことができる。
【０１０１】
他のいくつかの実施形態では、動作が、撮影シーンの一連の第２の時間間隔を決定することを含むことができる。一連の第２の時間間隔の各々は、一連の撮影シーンにおける撮

影シーンのオーディオ部分を再生するための時間間隔を示すことができる。動作は、撮影シーンのビデオ説明情報のオーディオ表現の第３の時間間隔を決定することをさらに含むことができる。動作は、決定された一連の第２の時間間隔及び決定された第３の時間間隔に基づいてビデオ説明情報のオーディオ表現を再生する速度を決定することをさらに含むことができる。動作は、決定された速度に基づいてビデオ説明情報のオーディオ表現の再生を第１の時間間隔において制御することをさらに含むことができる。第１の時間間隔は、一連の第２の時間間隔とは異なることができる。

【０１０２】

本開示の例示的な態様は、（回路２０２などの）回路を含むことができる（図１のメディアレンダリング装置１０２などの）メディアレンダリング装置を含むことができる。回路は、（一連の撮影シーン１１４などの）一連の撮影シーン、（オーディオ部分１１６などの）オーディオ部分及び（テキスト情報１１８などの）テキスト情報を含むことができるメディアコンテンツを検索するように構成することができる。テキスト情報は、（ビデオ説明情報１１８Ａなどの）テキストベースのビデオ説明情報及び（タイミング情報１１８Ｂなどの）タイミング情報を含むことができる。ビデオ説明情報１１８Ａは、一連の撮影シーンにおける（撮影シーン３０４などの）撮影シーンを説明することができる。メディアコンテンツは、一連の撮影シーンの各々のオーディオ部分を表すことができるクローズドキャプション情報をさらに含むことができる。一連の撮影シーンの各々を説明するビデオ説明情報は、クローズドキャプション情報と共にメディアコンテンツ内に符号化することができる。ある実施形態では、回路を、撮影シーンのテキスト情報をビデオ説明情報のオーディオ表現に変換するようにさらに構成することができる。

【０１０３】

ある実施形態では、回路を、撮影シーンのテキスト情報から、ビデオ説明情報を再生するためのタイミング情報を抽出するようにさらに構成することができる。回路は、ビデオ説明情報の再生を、抽出された撮影シーンのタイミング情報によって示される第１の時間間隔においてテキスト表現又はテキスト表現及びオーディオ表現のいずれかで制御するようにさらに構成することができる。

【０１０４】

別の実施形態では、回路を、撮影シーンのテキスト情報から、ビデオ説明情報を再生するための速度情報を抽出するようにさらに構成することができる。テキスト情報は、速度情報をさらに含むことができる。回路は、抽出された速度情報に基づいて、抽出された撮影シーンのタイミング情報によって示される第１の時間間隔においてビデオ説明情報のオーディオ表現の再生を制御するようにさらに構成することができる。

【０１０５】

いくつかの実施形態では、回路を、撮影シーンの（一連の第２の時間間隔４１６などの）一連の第２の時間間隔を決定するように構成することができる。一連の第２の時間間隔の各々は、一連の撮影シーンにおける撮影シーンのオーディオ部分を再生するための時間間隔を示すことができる。回路は、撮影シーンのビデオ説明情報のオーディオ表現の（第３の時間間隔４１８などの）第３の時間間隔を決定するようにさらに構成することができる。回路は、ビデオ説明情報のオーディオ表現を再生する速度を決定するようにさらに構成することができる。ビデオ説明情報のオーディオ表現を再生する速度は、決定された一連の第２の時間間隔及び決定された第３の時間間隔に基づいて決定することができる。ある実施形態では、決定される速度が、変換されたオーディオ表現の実際の再生速度よりも低いことができる。別の実施形態では、決定される速度が、変換されたオーディオ表現の実際の再生速度よりも高いことができる。

【０１０６】

いくつかの実施形態では、回路を、ビデオ説明情報のオーディオ表現の再生を（第１の時間間隔４２２などの）第１の時間間隔において制御するように構成することができる。ビデオ説明情報のオーディオ表現の再生は、決定された速度に基づいて制御することができる。ある実施形態では、回路を、撮影シーンの（一連の第４の時間間隔４２０Ａ～４２

０Ｃなどの）一連の第４の時間間隔を決定するように構成することができる。一連の第４の時間間隔の各々は、一連の第２の時間間隔とは異なることができる。回路は、一連の第４の時間間隔から、時間間隔閾値よりも高いことができる第１の時間間隔を選択するように構成することができる。第１の時間間隔は、一連の第２の時間間隔とは異なることができる。ある実施形態では、第１の時間間隔が、撮影シーンの第１のせりふと第２のせりふとの間であることができる。第１のせりふは、撮影シーンの第１のショットの最後の単語であることができ、第２のせりふは、撮影シーンの第２のショットの最初の単語であることができる。第１のショット及び第２のショットは、撮影シーンの連続するショットであることができる。別の実施形態では、第１の時間間隔が、撮影シーンの開始と撮影シーンの第１のせりふとの間であることができる。

【０１０７】

いくつかの実施形態では、回路を、メディアレンダリング装置に関連する規定の速度設定に基づいて、ビデオ説明情報のオーディオ表現を再生する速度を決定するように構成することができる。規定の速度設定は、ビデオ説明情報のオーディオ表現の最大再生速度を示すことができる。回路は、テキスト情報と共に速度情報を受け取り、決定された速度及び規定の速度設定に基づいて撮影シーンの画像部分又はオーディオ部分の一方の再生を制御するようにさらに構成することができる。いくつかの実施形態では、撮影シーンを説明するビデオ説明情報が、撮影シーン内に存在する生物オブジェクト又は無生物オブジェクトに関する認知情報を含むことができる。回路は、撮影シーンのビデオ説明情報に含まれる認知情報の再生を制御するように構成することができる。

【０１０８】

ある実施形態では、回路を、一連の撮影シーンのうちの１つの撮影シーンの説明に対応できる第１のユーザ入力を受け取るように構成することができる。回路は、受け取られた第１のユーザ入力を、一連の撮影シーンの各々に関連するビデオ説明情報内で検索するようにさらに構成することができる。回路は、検索に基づいて、メディアコンテンツを再生するための再生タイミング情報を決定するようにさらに構成することができる。回路は、決定された再生タイミング情報に基づいてメディアコンテンツの再生を制御するようにさらに構成することができる。

【０１０９】

別の実施形態では、回路を、メディアコンテンツをレンダリングできる対象であるユーザのプロファイル情報を示すことができる第２のユーザ入力を受け取るように構成することができる。回路は、受け取られた第２のユーザ入力に基づいて、ビデオ説明情報のオーディオ表現を再生する速度設定を決定するように構成することができる。

【０１１０】

いくつかの実施形態では、メディアレンダリング装置に関連する（メモリ２０４などの）メモリを、（訓練済み機械学習（ＭＬ）モデル２１２などの）訓練済みＭＬモデルを記憶するように構成することができる。回路は、撮影シーンの少なくとも１つの特性の分析に基づいて撮影シーンのコンテキスト情報を決定するように構成することができる。回路は、撮影シーンの決定されたコンテキスト情報に対する訓練済みＭＬモデルの適用に基づいて、ビデオ説明情報のオーディオ表現を再生するためのオーディオ特性を決定するようにさらに構成することができる。回路は、決定された速度及び決定されたオーディオ特性に基づいて、ビデオ説明情報のオーディオ表現の再生を第１の時間間隔において制御するようにさらに構成することができる。

【０１１１】

ある実施形態では、メディアレンダリング装置が、ビデオ説明情報のテキスト表現を再生する（又は表示する）ように構成されたディスプレイ装置を含むことができる。別の実施形態では、ビデオ説明情報のオーディオ表現の再生に加えてテキスト表現を表示することができる。

【０１１２】

別の実施形態では、回路を、オーディオレンダリング装置を制御するようにさらに構成

することができる。オーディオレンダリング装置は、メディアレンダリング装置に関連することができる。オーディオレンダリング装置は、ビデオ説明情報のオーディオ表現及び撮影シーンのオーディオ部分を再生するように制御することができる。

【０１１３】

本開示は、ハードウェアで実現することも、又はハードウェアとソフトウェアとの組み合わせで実現することもできる。本開示は、少なくとも１つのコンピュータシステム内で集中方式で実現することも、又は異なる要素を複数の相互接続されたコンピュータシステムにわたって分散できる分散方式で実現することもできる。本明細書で説明した方法を実行するように適合されたコンピュータシステム又はその他の装置が適することができる。ハードウェアとソフトウェアとの組み合わせは、ロードされて実行された時に本明細書で説明した方法を実行するようにコンピュータシステムを制御することができるコンピュータプログラムを含む汎用コンピュータシステムとすることができる。本開示は、他の機能も実行する集積回路の一部を含むハードウェアで実現することができる。

【０１１４】

本開示は、本明細書で説明した方法の実装を可能にする全ての特徴を含み、コンピュータシステムにロードされた時にこれらの方法を実行できるコンピュータプログラム製品に組み込むこともできる。本文脈におけるコンピュータプログラムとは、情報処理能力を有するシステムに特定の機能を直接的に、或いはａ）別の言語、コード又は表記法への変換、ｂ）異なる内容形態での複製、のいずれか又は両方を行った後に実行させるように意図された命令セットの、あらゆる言語、コード又は表記法におけるあらゆる表現を意味する。

【０１１５】

いくつかの実施形態を参照しながら本開示を説明したが、当業者であれば、本開示の範囲から逸脱することなく様々な変更を行うことができ、同等物を代用することもできると理解するであろう。また、本開示の範囲から逸脱することなく、特定の状況又は内容を本開示の教示に適合させるように多くの修正を行うこともできる。従って、本開示は、開示した特定の実施形態に限定されるものではなく、添付の特許請求の範囲内に収まる全ての実施形態を含むように意図される。

【符号の説明】

【０１１６】

１００　ネットワーク環境
１０２　メディアレンダリング装置
１０４　ディスプレイ装置
１０６　オーディオレンダリング装置
１０８　サーバ
１１０　通信ネットワーク
１１２　メディアコンテンツ
１１４　一連の撮影シーン
１１４Ａ　第１の撮影シーン
１１４Ｂ　第２の撮影シーン
１１４Ｎ　第Ｎの撮影シーン
１１６　オーディオ部分
１１８　テキスト情報
１１８Ａ　ビデオ説明情報
１１８Ｂ　タイミング情報
１１８Ｃ　速度情報
１２０　ユーザ

【図面】

【図１】

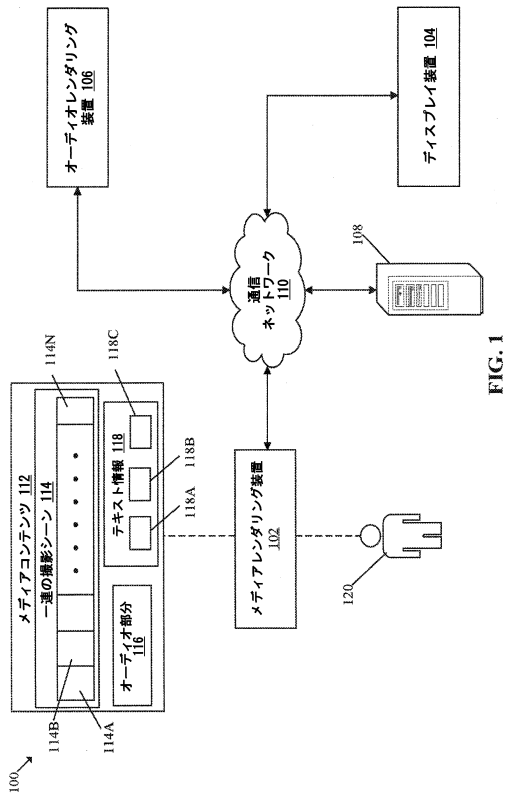

FIG. 1

【図２】

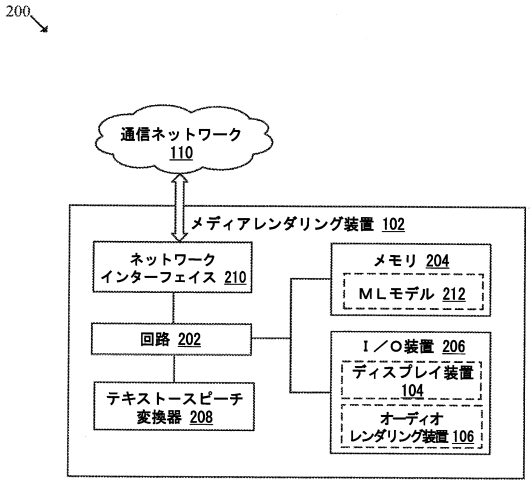

FIG. 2

【図３Ａ】

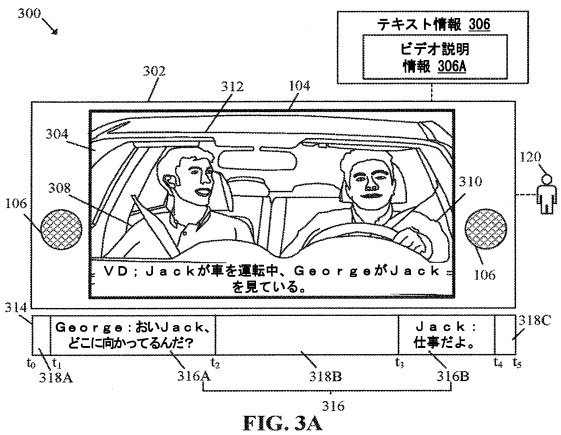

FIG. 3A

【図３Ｂ】

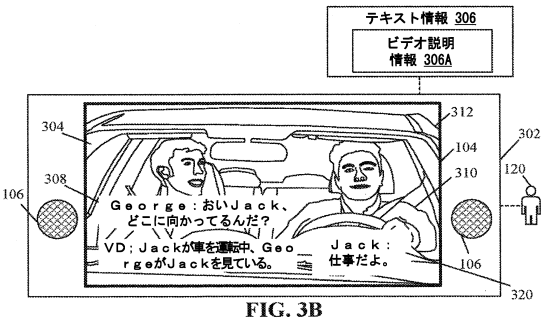

FIG. 3B

【図４】

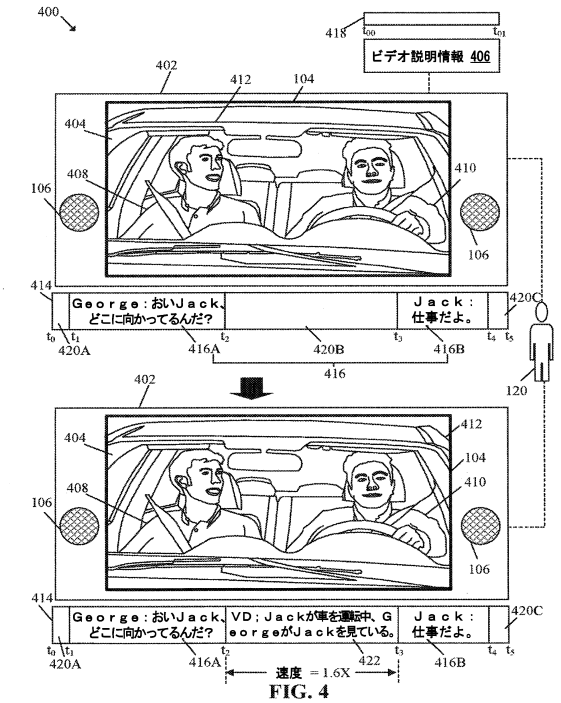

FIG. 4

【図５】

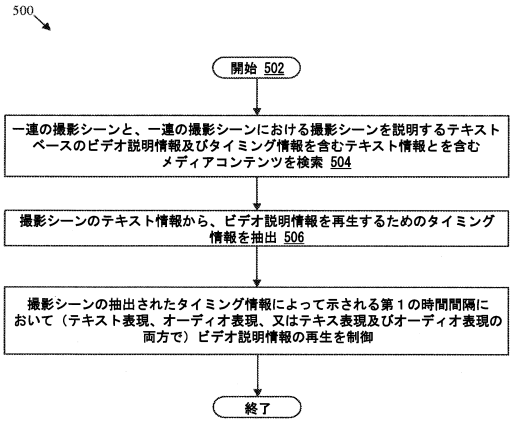

FIG. 5

【図６】

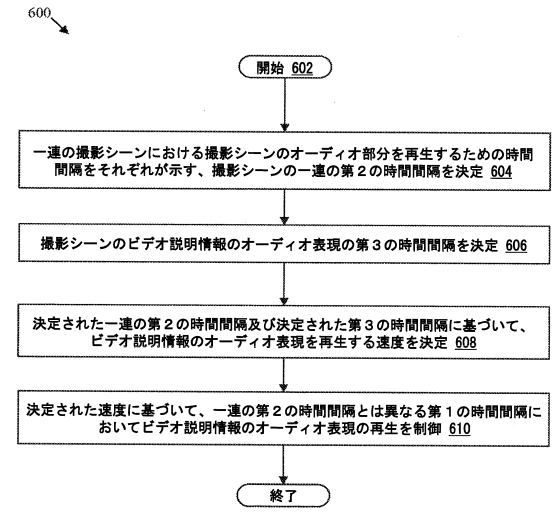

FIG. 6

フロントページの続き

１２７ サンディエゴ ヴィア エスプリロ １６５３５ ソニー コーポレイション オブ アメリカ内
(72)発明者 ネジャット マヤル マイク
アメリカ合衆国 カリフォルニア州 ９２１２７ サンディエゴ ヴィア エスプリロ １６５３５ ソニー コーポレイション オブ アメリカ内
(72)発明者 シンタニ ピーター
アメリカ合衆国 カリフォルニア州 ９２１２７ サンディエゴ ヴィア エスプリロ １６５３５ ソニー コーポレイション オブ アメリカ内
(72)発明者 ブランチャード ロバート
アメリカ合衆国 カリフォルニア州 ９２１２７ サンディエゴ ヴィア エスプリロ １６５３５ ソニー コーポレイション オブ アメリカ内

審査官 大西 宏

(56)参考文献
特開２０００－２５０５７５（ＪＰ，Ａ）
特開２００３－１４３５７５（ＪＰ，Ａ）
特開２００３－２５９３２０（ＪＰ，Ａ）
特開２００４－０６２７６９（ＪＰ，Ａ）
特表２０１７－５３１９３６（ＪＰ，Ａ）
国際公開第２０１８／２１１７４８（WO，Ａ１）
国際公開第２０２０／１１２８０８（WO，Ａ１）
Ruxandra Tapu et al.，DEEP-HEAR: A Multimodal Subtitle Positioning System Dedicated to Deaf and Hearing-Impaired People，IEEE Access，米国，IEEE，2019年07月01日，Volume: 7，pp.88150-88162，https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8751956，IEL Online

(58)調査した分野 (Int.Cl.，ＤＢ名)
Ｈ０４Ｎ ２１／００ －２１／８５８