

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第5555068号
(P5555068)

(45) 発行日 平成26年7月23日 (2014. 7. 23)

(24) 登録日 平成26年6月6日 (2014. 6. 6)

(51) Int. Cl. F I
HO 4 N 21/426 (2011. 01) HO 4 N 21/426
HO 4 N 21/434 (2011. 01) HO 4 N 21/434
HO 4 N 21/439 (2011. 01) HO 4 N 21/439

請求項の数 17 (全 16 頁)

(21) 出願番号	特願2010-137677 (P2010-137677)	(73) 特許権者	000001007
(22) 出願日	平成22年6月16日 (2010. 6. 16)		キヤノン株式会社
(65) 公開番号	特開2012-4835 (P2012-4835A)		東京都大田区下丸子3丁目30番2号
(43) 公開日	平成24年1月5日 (2012. 1. 5)	(74) 代理人	100076428
審査請求日	平成25年6月10日 (2013. 6. 10)		弁理士 大塚 康德
		(74) 代理人	100112508
			弁理士 高柳 司郎
		(74) 代理人	100115071
			弁理士 大塚 康弘
		(74) 代理人	100116894
			弁理士 木村 秀二
		(74) 代理人	100130409
			弁理士 下山 治
		(74) 代理人	100134175
			弁理士 永川 行光

最終頁に続く

(54) 【発明の名称】 再生装置及びその制御方法及びプログラム

(57) 【特許請求の範囲】

【請求項 1】

符号化階層を選択して復号することにより、前記符号化階層ごとの復号映像を得ることが可能な階層符号化映像データと、当該階層符号化映像データの所定の符号化階層に対応付けられた音声データと、を受信する受信手段と、

複数の符号化階層のうち、選択された1つの符号化階層を復号し、復号映像を取得する復号手段と、

前記複数の符号化階層のそれぞれを復号した場合に得られる復号映像を所定の空間解像度に変換した場合の面積を示す比較用面積の情報を取得する取得手段と、

前記復号手段によって復号される符号化階層に前記音声データが対応付けられていない場合に、前記音声データが対応付けられた符号化階層を復号した場合に得られる第1の復号映像の前記比較用面積の情報と、前記復号手段により復号される第2の復号映像の前記比較用面積の情報とに基づいて、前記受信手段により受信された前記音声データを補正する補正手段と、

を備えることを特徴とする映像処理装置。

【請求項 2】

前記補正手段は、前記復号手段によって復号される符号化階層に前記音声データが対応付けられていない場合に、前記第1の復号映像の比較用面積に対する前記第2の復号映像の比較用面積の比に基づいて、前記音声データを補正することを特徴とする請求項1に記載の映像処理装置。

10

20

【請求項 3】

前記補正手段は、前記第 1 の復号映像の比較用面積に対する前記第 2 の復号映像の比較用面積の比の平方根を算出し、

該平方根が 1 以上の場合は、予め定められた音量比率に該平方根を乗じて得られた音量比率を用いて前記音声データの音量を補正し、

該平方根が 1 未満の場合は、前記予め定められた音量比率に該平方根の負の逆数を乗じて得られた音量比率を用いて前記音声データの音量を補正する

ことを特徴とする請求項 1 または 2 に記載の映像処理装置。

【請求項 4】

前記階層符号化映像データのコンテンツのジャンルを識別する識別手段をさらに備え、

前記予め定められた音量比率は、ジャンルごとに定められおり、

前記補正手段は、前記識別手段により識別されたジャンルに対応する前記予め定められた音量比率を前記音声データの音量の補正に用いる

ことを特徴とする請求項 3 に記載の映像処理装置。

【請求項 5】

前記復号手段は、前記選択された 1 つの符号化階層の復号に、前記複数の符号化階層のうち前記選択された 1 つの符号化階層よりも下位の符号化階層の復号の結果を利用することを特徴とする請求項 1 乃至 4 のいずれか 1 項に記載の映像処理装置。

【請求項 6】

前記補正手段は、前記音声データの高音部、低音部、サラウンド、音像定位の少なくともいずれかを補正することを特徴とする請求項 1 乃至 5 のいずれか 1 項に記載の映像処理装置。

【請求項 7】

前記補正手段は、前記階層符号化映像データにそれぞれ異なる前記音声データが対応付けられた符号化階層が複数含まれる場合に、前記音声データが対応付けられた複数の符号化階層のうち、復号した場合に得られる復号映像の前記比較用面積が、前記第 2 の復号映像の比較用面積に近い符号化階層に対応付けられている前記音声データを補正することを特徴とする請求項 1 乃至 6 のいずれか 1 項に記載の映像処理装置。

【請求項 8】

前記音声データは、復号した場合に得られる復号映像の前記比較用面積が最も広い符号化階層に対応付けられていることを特徴とする請求項 1 乃至 7 のいずれか 1 項に記載の映像処理装置。

【請求項 9】

受信手段が、符号化階層を選択して復号することにより、前記符号化階層ごとの復号映像を得ることが可能な階層符号化映像データと、当該階層符号化映像データの所定の符号化階層に対応付けられた音声データと、を受信する受信工程と、

復号手段が、複数の符号化階層のうち、選択された 1 つの符号化階層を復号する復号工程と、

取得手段が、前記複数の符号化階層のそれぞれを復号した場合に得られる復号映像を所定の空間解像度に変換した場合の面積を示す比較用面積の情報を取得する取得工程と、

補正手段が、前記復号工程において復号される符号化階層に前記音声データが対応付けられていない場合に、前記音声データが対応付けられた符号化階層を復号した場合に得られる第 1 の復号映像の前記比較用面積の情報と、前記復号工程において復号される第 2 の復号映像の前記比較用面積の情報とに基づいて、前記受信工程において受信された前記音声データを補正する補正工程と、

を備えることを特徴とする映像処理装置の制御方法。

【請求項 10】

前記補正手段は前記補正工程において、前記復号工程において復号される符号化階層に前記音声データが対応付けられていない場合に、前記第 1 の復号映像の比較用面積に対する前記第 2 の復号映像の比較用面積の比に基づいて、前記音声データを補正することを特

10

20

30

40

50

徴とする請求項 9 に記載の映像処理装置の制御方法。

【請求項 1 1】

前記補正手段は前記補正工程において、前記第 1 の復号映像の比較用面積に対する前記第 2 の復号映像の比較用面積の比の平方根を算出し、

該平方根が 1 以上の場合は、予め定められた音量比率に該平方根を乗じて得られた音量比率を用いて前記音声データの音量を補正し、

該平方根が 1 未満の場合は、前記予め定められた音量比率に該平方根の負の逆数を乗じて得られた音量比率を用いて前記音声データの音量を補正することを特徴とする請求項 9 または 1 0 に記載の映像処理装置の制御方法。

【請求項 1 2】

識別手段が、前記階層符号化映像データのコンテンツのジャンルを識別する識別工程をさらに備え、

前記予め定められた音量比率は、ジャンルごとに定められており、

前記補正手段は前記補正工程において、前記識別工程において識別されたジャンルに対応する前記予め定められた音量比率を前記音声データの音量の補正に用いることを特徴とする請求項 1 1 に記載の映像処理装置の制御方法。

【請求項 1 3】

前記復号手段は前記復号工程において、前記選択された 1 つの符号化階層の復号に、前記複数の符号化階層のうち前記選択された 1 つの符号化階層よりも下位の符号化階層の復号の結果を利用することを特徴とする請求項 9 乃至 1 2 のいずれか 1 項に記載の映像処理装置の制御方法。

【請求項 1 4】

前記補正手段は前記補正工程において、前記音声データの高音部、低音部、サラウンド、音像定位の少なくともいずれかを補正することを特徴とする請求項 9 乃至 1 3 のいずれか 1 項に記載の映像処理装置の制御方法。

【請求項 1 5】

前記補正手段は前記補正工程において、前記階層符号化映像データにそれぞれ異なる前記音声データが対応付けられた符号化階層が複数含まれる場合に、前記音声データが対応付けられた複数の符号化階層のうち、復号した場合に得られる復号映像の前記比較用面積が、前記第 2 の復号映像の比較用面積に近い符号化階層に対応付けられている前記音声データを補正することを特徴とする請求項 9 乃至 1 4 のいずれか 1 項に記載の映像処理装置の制御方法。

【請求項 1 6】

前記音声データは、復号した場合に得られる復号映像の前記比較用面積が最も広い符号化階層に対応付けられていることを特徴とする請求項 9 乃至 1 5 のいずれか 1 項に記載の映像処理装置の制御方法。

【請求項 1 7】

コンピュータを、請求項 1 乃至 8 のいずれか 1 項に記載の映像処理装置の各手段として機能させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【0 0 0 1】

本発明は、一つの映像ストリームデータ中に複数の解像度や画角の映像が階層符号化されている階層符号化映像データと、及び当該階層符号化映像データのある符号化階層に対応する音声データとの再生技術に関する。

【背景技術】

【0 0 0 2】

近年、地上デジタル放送において、MPEG2 - Video 符号化方式にて符号化された解像度 720 × 480 ピクセルや 1440 × 1080 ピクセルの映像が放送されている。地上デジタル放送は、携帯電話などの携帯機器向けにもワンセグメント放送として、H

10

20

30

40

50

、264/AVC(Audio Visual Coding)符号化方式にて符号化された320×240ピクセルの映像が別ストリームで放送されている。

【0003】

一方、H.264/AVCの拡張として、映像ストリームデータ中に複数の解像度の映像を符号化可能な、H.264/SVC(Scalable Video Coding)技術が規格化されている。H.264/SVC規格では、例えば320×240、720×480、1440×1080ピクセル等の解像度が異なる複数の映像を、異なる符号化階層(レイヤとも呼ぶ)として1つの映像ストリームデータの中に階層符号化している。このように異なる解像度の映像を1つのストリームデータの中に符号化することで、別々の映像ストリームを伝送する場合に比べて効率よく圧縮・伝送することができる。

10

【0004】

またH.264/SVC規格では、画角が異なる複数の映像を、1つの映像ストリームデータの中に符号化することも可能である。例えば、サッカーグラウンド全体を映した全景映像と、当該全景映像中のサッカー選手の領域だけを映した特定領域映像とが、それぞれ異なるレイヤに階層符号化される。そして再生時にレイヤを選択的に復号することで、視聴中の映像の画角変更や、表示装置の表示解像度及び表示アスペクト比に適した映像を再生することが可能となる。

【0005】

このように、H.264/SVC規格を用いることで、地上デジタル放送とワンセグメント放送のように異なるストリームで映像を伝送することなく、1つの映像ストリームデータの伝送で複数の種類の表示装置に対応可能となる。即ち、伝送帯域の効率化や、複数の映像サイズや画角がユーザによって選択可能とするサービス提供が図れるため、将来H.264/SVC規格のような階層符号化映像データがテレビ放送に利用されることが想定される。

20

【0006】

なお、テレビ放送に階層符号化映像データを利用する場合であっても、音声ストリームデータは現在の放送と同様に、1つの音声ストリームデータのみを有することが考えられる。上述したように、階層符号化映像データを用いることによりユーザはレイヤを選択して視聴する映像の画角を変更することが可能であるが、ある1つの符号化階層に対応付けられた音声ストリームデータのみが提供される場合は次のような問題が生じる。即ち、映像の画角が変更されても再生される音声ストリームデータは変化しないため、画角変更によってユーザに臨場感を与える効果が低いという問題である。

30

【0007】

特許文献1には、映像の一部を拡大することにより画角の変化が生じた場合に、映像の切り出しサイズや位置に応じて音声の音像定位を変更することにより、ユーザに臨場感を与える技術が開示されている。

【先行技術文献】

【特許文献】

【0008】

【特許文献1】特開平2004-336430公報

40

【発明の概要】

【発明が解決しようとする課題】

【0009】

しかしながら、特許文献1では、階層符号化映像データのような、複数の画角の映像が階層符号化された映像データの再生について考慮されていなかった。複数の画角の映像が階層符号化された映像ストリームデータに対し、1つのレイヤ(即ち、1つの画角)に対応した音声ストリームデータのみが与えられる場合、従来のように映像と音声とが1対1に対応して与えられる場合とは異なる処理が必要となる。例えば、サッカー中継の映像コンテンツにおいて、グラウンド全体の映像と、グラウンドの一部注目領域映像が映像ストリームデータに階層符号化されており、1つの音声ストリームデータが一部注目領域の映像

50

のレイヤに対応している場合を想定する。このような場合に従来技術のようにグランド全体の映像のレイヤに音声ストリームが対応しているものとして処理してしまうと、一部注目領域の映像が選択された際に、必要のない音声補正処理が適用されてしまう。

【 0 0 1 0 】

また、階層符号化された映像ストリームデータには、解像度は同一で画角が異なる複数の映像が含まれる場合もあるが、特許文献 1 ではこのような場合について考慮されていない。

【 0 0 1 1 】

本発明は、上述の問題点に鑑みてなされたものである。即ち、階層符号化映像データと、階層符号化映像データの特定の符号化階層に対応付けられた音声データとの再生において、音声データが対応付けられていない符号化階層の映像を再生する場合でもユーザに臨場感を与える音声再生を行うことを目的とする。

【課題を解決するための手段】

【 0 0 1 2 】

前述の目的を達成するために、本発明の映像処理装置は、以下の構成を備える。

符号化階層を選択して復号することにより、符号化階層ごとの復号映像を得ることが可能な階層符号化映像データと、当該階層符号化映像データの所定の符号化階層に対応付けられた音声データと、を受信する受信手段と、複数の符号化階層のうち、音声データが対応付けられている音声対応符号化階層を特定する特定手段と、複数の符号化階層のうち、選択された 1 つの符号化階層を復号し、復号映像を取得する復号手段と、複数の符号化階層のそれぞれを復号した場合に得られる復号映像を所定の空間解像度に変換した場合の面積を示す比較用面積の情報を取得する取得手段と、復号手段によって復号される符号化階層に音声データが対応付けられていない場合、音声データが対応付けられた符号化階層を復号した場合に得られる第 1 の復号映像の比較用面積の情報と、復号手段により復号される第 2 の復号映像の比較用面積の情報とに基づいて、受信手段により受信された音声データを補正する補正手段と、を備えることを特徴とする。

【発明の効果】

【 0 0 1 3 】

このような構成により本発明は次の効果を奏する。階層符号化映像データと、階層符号化映像データの特定の符号化階層に対応付けられた音声データとの再生において、音声データが対応付けられていない符号化階層の映像を再生する場合でもユーザに臨場感を与える音声再生を行うことを可能とする。

【図面の簡単な説明】

【 0 0 1 4 】

【図 1】実施形態に係るセットトップボックスの機能構成を示したブロック図。

【図 2】セットトップボックスの音声補正制御部の機能構成を示すブロック図。

【図 3】コンテンツデータの構成を説明するための図。

【図 4】コンポーネント記述子を説明するための図。

【図 5】付加情報データを説明するための図。

【図 6】実施形態における音声補正制御処理のフローチャート。

【図 7】音声補正量情報テーブルの構成例を示した図。

【図 8】階層符号化映像データの各符号化階層が表現する復号映像を説明するための図。

【発明を実施するための形態】

【 0 0 1 5 】

以下、本発明の好適かつ例示的な一実施形態について、図面を参照して詳細に説明する。なお、以下に説明する一実施形態は、再生装置の一例としての、階層符号化映像データと当該階層符号化映像データの 1 つの符号化階層に対応付けられた音声データとを再生可能なセットトップボックスに、本発明を適用した例を説明する。しかし、本発明は、階層符号化映像データと、当該階層符号化映像データの所定の符号化階層に対応付けられた音声データとを再生可能な任意の機器に適用可能である。

【 0 0 1 6 】

本実施形態では、後述するように、再生装置で再生される際に初期状態で選択される階層符号化映像データの符号化階層（default_layer）が予め定められている。音声データは、階層符号化映像データの複数の符号化階層のうち、当該初期状態で選択される符号化階層に対応づけられているものとして説明する。上述の通り、符号化階層ごとに映像が符号化されており、ある符号化階層を復号することにより、対応する解像度と画角の復号映像が得られる。

【 0 0 1 7 】

なお、以下の説明では階層符号化方法の一例としてH.264/SVC規格を想定している。当該規格では、ある符号化階層を復号する場合、それより下位の符号化階層のデータを利用する場合がある。以下の説明において、ある符号化階層を最上位階層として復号するとの記載はこのような場合を想定したものである。しかしながら、本明細書において「ある符号化階層を復号する」とは、最終的にその符号化階層の復号映像を得ることであり、復号の過程で他の符号化階層のデータを用いるかどうかは本発明とは無関係である。従って、本発明は特定の階層符号化方法に限定されないことに留意されたい。

【 0 0 1 8 】

なお本実施形態において、音声データは、階層符号化映像データのある符号化階層に対応付けられている。そして音声データは、当該音声データが対応付けられている符号化階層を復号して得られる映像に対応した位置で収録されているものとする。つまり、被写体までの距離が50mの位置で撮影された等倍映像に対して音声データが対応付けられている場合は、その音声データは、映像の撮影時、人間がその撮影位置（被写体までの距離が50m）において実際に聞こえる音声に相当するものである。なお、ある位置で撮影した等倍映像とは、その位置における人間の視野に等しい映像である。例えば、カメラが望遠レンズなど、人間と異なる画角を有するレンズで撮影する場合、撮影された映像を人間の画角に等しい画角のレンズで撮影するためのカメラ位置（換算撮影位置）が、望ましい音声データ音声収録位置である。このように、本実施形態においては、被写体から100m離れた位置で収録された音声データが、被写体までの距離が10mに相当するように撮影された復号映像に対応付けられることはないことを想定している。なお、撮影位置（換算撮影位置を含む）と音声収録位置は同一位置であることが望ましいが、本実施形態では音声収録位置が撮影位置と異なっても、被写体までの距離が同一であれば、撮影位置で聞こえる音声とほぼ同じ音声の収録できるものとして説明を行う。

【 0 0 1 9 】

なお、以下の説明では、音声データが対応付けられている符号化階層を音声対応符号化階層と称し、音声データが対応づけられていない符号化階層と区別する。また、以下の説明では、階層符号化映像データ、音声データ及び付加情報データを含んだデータストリームをコンテンツデータと呼ぶ。コンテンツデータからは、符号化階層それぞれについて得られる復号映像の画角や解像度の情報が取得できるものとする。

【 0 0 2 0 】

図1は、本発明の実施形態に係るセットトップボックスの機能構成を示すブロック図である。セットトップボックス100は、再生制御部101、コンテンツ取得部102、コンテンツ解析部103、映像処理部104、音声処理部105、操作受信部106、音声補正制御部107、及び音声出力調整部108を備える。セットトップボックス100には、電波及びネットワークで伝送されるコンテンツデータが、アンテナ入力やネットワーク入力より入力される。

【 0 0 2 1 】

コンテンツ取得部102は、再生制御部101からの指示により、電波及びネットワークで伝送された信号に復調処理及び誤り訂正処理を適用し、受信したコンテンツデータを、コンテンツ解析部103に出力する。

【 0 0 2 2 】

コンテンツ解析部103は、再生制御部101からの指示により、コンテンツ取得部1

10

20

30

40

50

02から入力されたコンテンツデータを解析する。そして、コンテンツ解析部103はコンテンツデータ内に多重化されている階層符号化映像データ、音声データ、及び付加情報データを分離する。コンテンツ解析部103は、分離した階層符号化映像データを映像処理部104に、音声データを音声処理部105に、付加情報データを再生制御部101に出力する。

【0023】

映像処理部104は、再生制御部101からの再生指示により、再生指示された符号化階層を最上位階層として、階層符号化映像データの複数の符号化階層を選択的に復号し（映像復号）、得られた復号映像を外部接続された表示装置（ディスプレイ）に出力する。

【0024】

音声処理部105は、再生制御部101からの再生指示により、コンテンツ解析部103で分離され、音声処理部105に入力された音声データを復号し（音声復号）、得られた音声を音声出力調整部108に出力する。

【0025】

音声出力調整部108は、音声補正制御部107からの音声出力制御指示により、音声処理部105で出力された音声のサラウンドや低音高音などの音質調整や、LRなど各音声出力チャンネルの音像定位の補正処理を行う。そして音声出力調整部108は補正した音声を、外部接続された音声出力装置（スピーカ）に出力する。

【0026】

操作受信部106は、例えば赤外線受信ユニットであり、ユーザによって操作された不図示のリモートコントローラからのリモートコントローラ信号（キーコードデータ）等のユーザ操作信号を受信する。そして操作受信部106は、受信したユーザ操作信号を解析して操作された内容を再生制御部101に出力する。具体的には、操作受信部106は、ユーザ操作信号を、例えば再生開始、階層符号化映像データの各階層選択情報などの再生制御に関する信号に変換して、再生制御部101へ出力する。

【0027】

再生制御部101は、操作受信部106から入力された再生制御信号に応じて、コンテンツ取得部102へコンテンツデータ受信指示を行う。さらに再生制御部101は、コンテンツ解析部103へ、コンテンツ取得部102から入力されるコンテンツデータの解析指示を行い、コンテンツ解析部103が解析した付加情報データを取得する。また、再生制御部101は、コンテンツ解析部103から入力された付加情報データと操作受信部106から入力された再生制御信号に応じて、復号する階層符号化映像データの符号化階層、及び音声データを決定する。そして再生制御部101は、コンテンツ解析部103へ再生を行う復号映像及び音声をそれぞれ映像処理部104、音声処理部105に指示する。その後再生制御部101は、映像処理部104へ決定した復号する符号化階層の復号映像の再生指示を行う。

【0028】

また再生制御部101は、コンテンツ解析部103から入力される付加情報データと操作受信部106から出力された再生制御信号に応じて、音声処理部105へ音声データの復号及び再生指示を行う。再生制御部101は、コンテンツ解析部103から入力される付加情報データと操作受信部106から出力された再生制御信号に応じて、音声補正制御部107に付加情報データとあわせて音声補正指示を出力する。

【0029】

音声補正制御部107は、再生制御部101から入力された付加情報データと音声補正指示に応じて、音声再生に係るサラウンドや低音高音等の音質調整や、各音声出力チャンネルの音像定位の補正量を決定し、音声出力調整部108へ音声出力調整の指示を行う。

【0030】

ここで図2を用いて、セットトップボックス100における音声補正制御部107で行う処理をさらに詳細に説明する。音声補正制御部107は、映像音声対応判定部201、再生映像画角比判定部202、音声補正量決定部203、及び音声補正情報保持部204

10

20

30

40

50

を備える。

【 0 0 3 1 】

映像音声対応判定部 2 0 1 には、再生制御部 1 0 1 から再生状態、及び階層符号化映像データの各符号化階層の構造情報を含む付加情報データ等の、再生制御に関する情報が入力される。映像音声対応判定部 2 0 1 は、入力された各符号化階層の構造情報を元に、音声データがどの符号化階層に対応付けられているかの特定を行う。そして映像音声対応判定部 2 0 1 は、再生映像画角比判定部 2 0 2 へ特定した音声対応符号化階層、及び再生付加情報データ内の各符号化階層の情報を出力する。

【 0 0 3 2 】

再生映像画角比判定部 2 0 2 は、音声対応符号化階層の復号映像と、再生を行う符号化階層の復号映像との画角の比を算出し、その算出結果を音声補正量決定部 2 0 3 へ出力する。

10

【 0 0 3 3 】

音声補正量決定部 2 0 3 は、再生映像画角比判定部 2 0 2 で算出された画角の比を用いて、音声出力調整部 1 0 8 に指示する音声データの音声補正量を決定する。具体的には、音声補正量決定部 2 0 3 は、コンテンツデータのジャンルに応じた音声補正情報テーブルを保持している音声補正情報保持部 2 0 4 を参照し、受信しているコンテンツデータのジャンルに対応する音声補正量を取得する。そして音声補正量決定部 2 0 3 は、音声補正情報保持部 2 0 4 から取得した音声補正量に画角の比を乗じて得られた音声補正量を、音声出力調整部 1 0 8 に出力する。

20

【 0 0 3 4 】

図 3 は、図 1 に示したコンテンツ取得部 1 0 2 から出力される、階層符号化映像データを含んだコンテンツデータの構成例を示す図である。ここでは、I E C (International Electro-technical Commission、国際電気標準会議)、I E E E (Institute of Electrical and Electronic Engineers、電気電子学会等により規定されている T S 信号として構成される例を示している。

【 0 0 3 5 】

図 3 に示すように、コンテンツ取得部 1 0 2 が受信する信号には、複数の T S パケットが時分割多重されて T S 信号が構成されている。階層符号化映像データの v i d e o 部、音声データの a u d i o 部、及び付加情報データの d a t a 部は、それぞれ T S パケット単位で受信される。コンテンツ解析部 1 0 3 は、このような T S 信号データの解析を行い、それぞれ v i d e o 部、a u d i o 部、及び d a t a 部に分離し、v i d e o 部を映像処理部 1 0 4 に、a u d i o 部を音声処理部 1 0 5 に出力する。またコンテンツ解析部 1 0 3 は、v i d e o 部、a u d i o 部、及び d a t a 部の解析結果を再生制御部 1 0 1 に出力する。

30

【 0 0 3 6 】

ここで、複数の d a t a 部を収集して再構成される管理情報テーブルは、P A T (Program Association Table)、P M T (Program Map Table)、E I T (Event Information Table)、N I T (Network Information Table) 等で構成される。E I T は受信中の番組及びその後の番組編成を示すものであり、最初の部分に、テーブル I D、サービス I D、イベント I D、番組の放送開始時刻、放送継続時間等が記述された後に、いくつかの記述子が配置される。

40

【 0 0 3 7 】

コンテンツ解析部 1 0 3 では、この E I T の情報を元にどのような番組を受信しているのかの解析を行う。ここで、この E I T 情報を元に、受信しているコンテンツデータが表示領域変更可能番組であるかどうか、及びどのような表示領域選択が可能であるかの判定に関係がある、基本データ(フィールド)情報、及び各種記述子について説明する。

【 0 0 3 8 】

まずテーブル I D は、情報が E I T であることを識別する情報である。またテーブル I D に続いて記述されているイベント I D より、番組を識別することが可能である。また、

50

その番組毎に開始時刻とその番組の放送継続時間が記述されており、その開始時間 + 継続時間から、いつ番組が終了するかを判定可能である。

【 0 0 3 9 】

また E I T に含まれる記述子について、以下に説明する。記述子としては、例えば映像の解像度及びアスペクトの情報等を示すコンポーネント記述子、番組名を示す単形式イベント記述子、番組の続きがどのサービスで放送されるかといったイベントリレー情報が記述されたイベントグループ記述子等がある。図 3 に示したコンテンツ記述子には、番組のジャンルとして、「報道」、「スポーツ」、「ドラマ」、「映画」、「バラエティ」、「教養」等が記述されている。

【 0 0 4 0 】

図 4 は、上述したコンポーネント記述子を用いて階層符号化映像データの符号化階層情報を記述した付加情報データの例である。

コンポーネント記述子には階層符号化映像データの情報が記述されており、例えば component_type 識別子には、0 x 0 1 の場合は 4 8 0 i、0 x B 1 の場合は 1 0 8 0 i といった解像度情報などが記述されている。またコンポーネント記述子内には、コンテンツデータの再生開始時に初期状態で、どの符号化階層を選択して復号するかを示した情報として、図 4 に下線で示した default_layer 識別子が含まれる。default_layer 識別子の値が 1 であれば最初の符号化階層、即ちベースレイヤが初期状態で選択される符号化階層となる。また、default_layer 識別子の値が 2 であれば、2 番目の符号化階層、即ち 1 番目のエンハンスレイヤが初期状態で選択される符号化階層となる。即ち、default_layer 識別子において階層 ID を指定することで、複数のレイヤ構造、及び複数の画角が階層符号化映像データ内にある場合に、初期表示する符号化階層を決定することが可能となる。

【 0 0 4 1 】

ここで、階層符号化された映像の構成について、図 5 を用いて説明する。図 5 は、H. 2 6 4 / S V C 規格の階層符号化映像データの符号化階層の情報の例を示している。符号化階層の情報は、全体の符号化に関する情報を持つ S P S (SequenceParameterSet)、映像の符号化に関する P P S (PictureParameterSet)、実際の映像データである A U (AccessUnit) で構成され、それぞれの情報はさらに階層毎の情報に分かれている。

【 0 0 4 2 】

S P S 内には、各符号化階層の映像解像度情報 video_format に加え、各符号化階層間のオフセット情報として top_offset、left_offset、right_offset、bottom_offset というオフセット情報が記述されている。即ちこのオフセット情報より、各符号化階層間の画角の違いを判定することが可能であり、例えば各オフセット値が 1 0 であれば、基準となるレイヤに比べて上下左右に 1 0 ピクセル大きい画角を有する復号映像が得られると判定することができる。なお、オフセット情報とは各符号化階層の復号映像を同一の空間解像度にした場合に、1 つの符号化階層の復号映像で表現されている被写体の領域が、他の符号化階層の復号映像において表示される位置を表した値である。即ちオフセット情報は、1 つの符号化階層の復号映像で表現されている被写体の領域が、他の符号化階層の復号映像において表示される位置を、当該他の符号化階層の復号映像の 4 辺からのピクセル数で表した値である。

【 0 0 4 3 】

(音声補正制御処理)

以下、上述のような構成をもつ本実施形態のセットトップボックス 1 0 0 の音声補正制御処理について、図 6 のフローチャートをさらに用いて、具体的な処理を説明する。当該フローチャートに対応する処理は、再生制御部 1 0 1 が、例えば不図示の不揮発性メモリに記憶されている対応する処理プログラムを読み出し、不図示の R A M に展開して実行することにより実現することができる。なお、本音声補正制御処理は、例えばユーザによって操作されたりモートコントローラからコンテンツデータの再生開始要求を、操作受信部 1 0 6 が受信した際に開始されればよく、コンテンツデータの再生中は繰り返し実行されるものとして説明する。具体的には、操作受信部 1 0 6 よりユーザからのコンテンツデー

10

20

30

40

50

タの再生要求が入力されると、再生制御部 101 はセットトップボックス 100 の各ブロックにコンテンツデータの再生指示を行って再生処理を開始するとともに、本音声補正制御処理を開始する。

【0044】

S601 で、再生制御部 101 は音声補正制御部 107 に、受信しているコンテンツデータの付加情報データに含まれる番組情報を入力し、現在受信しているコンテンツデータが有する音声データが対応付けられている音声対応符号化階層を特定させる。具体的には、音声補正制御部 107 の映像音声対応判定部 201 は、コンポーネント記述子に記述されている default_layer の情報から、初期状態で表示される符号化階層に音声データが対応付けられていると特定する。

10

【0045】

S602 で、再生制御部 101 は、音声補正制御部 107 が特定した、音声対応符号化階層を復号して得られる復号映像と、現在選択されている符号化階層を復号して得られる復号映像の画角が異なるか否かを判定する。具体的には再生制御部 101 は、コンテンツ解析部 103 から受信したコンテンツデータが有する付加情報データのうち、SPS の各階層間のオフセット情報を参照して判定する。例えば、音声対応符号化階層と、現在選択されている符号化階層の間のオフセット情報 left_offset、right_offset、top_offset、及び bottom_offset の少なくとも 1 つが 0 以外の値をとる場合、再生制御部 101 は画角が変化していると判定する。再生制御部 101 は、音声対応符号化階層の復号映像と、現在選択されている符号化階層の復号映像の画角が異なる場合は処理を S603 に移し、同一である場合は処理を S605 に移す。

20

【0046】

S603 で、再生制御部 101 は音声補正制御部 107 に、階層符号化映像データの有する複数の符号化階層のうち、再生を行う符号化階層と、初期状態で表示される符号化階層との、それぞれの符号化階層を復号して得られる復号映像の画角の比を算出させる。具体的には、再生制御部 101 は音声対応符号化階層を復号して得られる復号映像と、再生を行う符号化階層を復号して得られる復号映像との画角の比を、音声補正制御部 107 に算出させる。

【0047】

まず音声補正制御部 107 の再生映像画角比判定部 202 は、音声対応符号化階層、及び再生を行う符号化階層の付加情報データから、それぞれの符号化階層を復号した際に得られる復号映像の画角を取得する。そして再生映像画角比判定部 202 は、音声対応符号化階層と、再生を行う符号化階層の画角の情報から、画角の比を算出する。なお、本実施形態では画角の比は、復号映像の解像度の比ではなく、それぞれの復号映像を同一の空間解像度に変換した場合の表示領域面積の比率の平方根で規定される。また音声の調整に用いるため、画角の比が 1 未満の数字となる場合は、画角の比の負の逆数を画角の比として規定する。例えば算出された画角の比が $1/2$ 倍である場合は、「 $-1/(1/2) = -2$ 」を画角の比として、再生映像画角比判定部 202 は音声補正量決定部 203 に出力する。

30

【0048】

ここで図 8 を用いて、具体的な画角の比の算出方法について例示する。図 8 の例では、H.264/AVC 規格と互換性をもつベースレイヤと、当該ベースレイヤの拡張である上位層のエンハンスレイヤを 2 つ有する階層符号化映像データの、各符号化階層を復号した場合の復号映像の位置関係を表している。なお、本実施形態では、各符号化階層を復号して得られる復号映像は、全て同一の空間解像度を有するものとする。

40

【0049】

まず、コンテンツデータが有する付加情報データのうち、SPS の各階層間のオフセット情報を参照する。コンポーネント記述子により default_layer がエンハンスレイヤ 1、現在再生を行うとして選択されている符号化階層がエンハンスレイヤ 2 である場合に、エンハンスレイヤ 1 を基準としたエンハンスレイヤ 2 のオフセット情報が以下であるとする

50

。

left_offset : 480
 right_offset : 480
 top_offset : 270
 bottom_offset : 270

【0050】

このとき、エンハンスレイヤ1の復号映像の表示解像度は960×540であるから、基準とするエンハンスレイヤ1に対するエンハンスレイヤ2の復号映像の縦横の解像度比は、

【0051】

【数1】

$$\frac{\text{エンハンスレイヤ1の水平解像度} + \text{left_offset} + \text{right_offset}}{\text{エンハンスレイヤ1の水平解像度}} = \frac{960 + 480 + 480}{960} = 2$$

$$\frac{\text{エンハンスレイヤ1の垂直解像度} + \text{top_offset} + \text{bottom_offset}}{\text{エンハンスレイヤ1の垂直解像度}} = \frac{540 + 270 + 270}{540} = 2$$

であり、表示領域面積の比は2×2=4倍となる。このため、エンハンスレイヤ1に対するエンハンスレイヤ2の画角の比は平方根をとり、2倍となる。

【0052】

S604で、再生制御部101は、受信しているコンテンツデータのジャンルと、現在選択されている符号化階層の復号映像とに適した音声補正量を音声補正制御部107に決定させる。具体的には、まず音声補正制御部107の音声補正量決定部203は音声補正情報保持部204が保持している音声補正量情報テーブルから、受信しているコンテンツデータのジャンルに対応した音声補正量を取得する。コンテンツデータのジャンルの情報は、受信したコンテンツデータが有する付加情報データに含まれるコンテンツ記述子から取得する。図7は音声補正量情報テーブルの例であり、音声補正情報保持部204は、音声補正情報テーブルとして、高音、低音、サラウンド（音の拡がり感）の音声補正量を、コンテンツデータのジャンル毎に保持している。そして、音声補正制御部107の音声補正量決定部203は、S603で算出した画角の比を、受信しているコンテンツデータに対応した音声補正量のそれぞれのパラメータに乗ずることにより、画角の違いを反映させた音声補正量を決定する。

【0053】

例えば図8に示すように、コンテンツデータがサッカー中継（ジャンルが「スポーツ」）であり、音声対応符号化階層がエンハンスレイヤ1、現在選択されている再生する符号化階層がエンハンスレイヤ2である場合について考える。

【0054】

エンハンスレイヤ2はエンハンスレイヤ1より広い画角であり、グラウンドを遠方から視聴することになるため、エンハンスレイヤ2の復号映像に対応する音声は、エンハンスレイヤ1のグラウンド中央の音声に比べ、臨場感の効果は小さくなる。具体的には、グラウンド内でボールが弾んだ音等の低音部は音量が小さくなり、観客やグラウンド全体に響く高音部の音量が大きくなるとともに、音の拡がり感が向上するように音声を補正する処理を行う。即ち、音声補正量情報テーブルから取得した、各調整パラメータの音声補正量に、S603で算出された画角の比（2倍）を乗じて、

高音部 : 20% × 2倍 = 40%
 低音部 : -20% × 2倍 = -40%
 サラウンド : 20% × 2倍 = 40%

が、最終的に決定された音声補正量となる。

【0055】

音声対応符号化階層の復号映像と、現在選択されている符号化階層の復号映像との間に画角の差がなかった場合も、S605で再生制御部101は、コンテンツデータのジャン

10

20

30

40

50

ルに適した音声補正用を音声補正制御部 107 に決定させる。なお、本ステップでは画角の違いを反映させた音声補正を行う必要がないため、音声補正制御部 107 の音声補正量決定部 203 は、音声補正情報保持部 204 より取得したコンテンツデータのジャンルに対応した音声補正量を、そのまま音声補正量とすればよい。

【0056】

そして S606 で、再生制御部 101 は、音声補正制御部 107 に決定させた音声補正量を用いて、音声処理部 105 で復号された音声について、音声出力調整部 108 に補正を適用させる。そして補正が適用された音声は、音声出力調整部 108 からセットトップボックス 100 に接続されている、スピーカ等の音声再生装置から出力される。

【0057】

なお、本実施形態ではコンテンツデータのジャンルに対応した音声補正量に対し、音声対応符号化階層の復号映像の画角に対する、選択されている符号化階層の復号映像の画角の比を乗じて得られた音声補正量を用いた、音声の補正方法について説明した。しかしながら、本発明はコンテンツデータのジャンルに対応した音声補正量を用いず、単に音声に対して画角の比を乗ずることにより、画角の違いによる音声の臨場感を作り出してもよい。

【0058】

また、上述した実施形態では、音声データは、その音声データが対応付けられている符号化階層を復号して得られる復号映像に対応した位置で収録されているものとして説明した。しかし、本発明において、音声データの収録位置から被写体までの距離と、当該音声データが対応付けられている符号化階層を復号して得られる映像に対応した撮影位置から被写体までの距離とが異なっている場合にも対応可能である。つまり、上述した実施形態の処理の前処理として、撮影位置と音声収録位置が異なる場合に、あたかも撮影位置で収録された音声であるかのように音の定位を補正する補正処理を実行すれば良い。この補正処理により、まずは、撮影位置で収録された音声であるように補正した上で、その後、補正した音声データに対して上述した実施形態に記載の制御を適用すれば良い。なお、音の定位や音像の補正技術は、ここに挙げるまでもなく古くから多数存在しているため、本発明を実施する上で好適な補正技術を適用すればよい。

【0059】

また、上述した実施形態では音声の高音、低音、サラウンドについて調整を行ったが、各映像階層のオフセット値を元に音像定位のずれを判定し、そのずれに応じて音像定位を補正することも可能である。

【0060】

また、上述した実施形態においては、理解及び説明を容易にするため、コンテンツデータが 1 つの符号化階層に対応付けられた 1 つの音声データを含む場合について説明した。しかし、本発明は音声データが 1 つである場合に限定されるものではなく、再生する符号化階層に対応付けられた音声データがコンテンツデータによって提供されない場合に適用可能である。例えば、音声データが 2 つ提供されている場合、再生する符号化階層の復号映像の画角に近い画角を有する符号化階層に対応付けられている音声データを用いることができる。

【0061】

また、上述した実施形態において、音声データが対応付けられた音声符号化階層は、コンテンツデータが有する付加情報データに記述されている、初期状態で再生される符号化階層であるものとして説明した。しかし、本発明の実施は付加情報データに初期状態で再生される符号化階層が記述されている場合に限定されるものではなく、例えば階層符号化映像データの各符号化階層の情報から決定してもよい。例えば、全景映像と、当該全景映像のうちの所定の領域を切り出した映像とが階層符号化されているようなコンテンツデータである場合、当該全景映像を符号化している符号化階層を、音声対応符号化階層として決定してもよい。即ち、階層符号化映像データの複数の符号化階層のうち、最も画角の広い復号映像が得られる符号化階層を、音声対応符号化階層としてもよい。つまり、1 つの

10

20

30

40

50

撮影位置で撮像された全景映像から階層符号化映像データが構成されているため、このようなコンテンツデータは当該1つの撮影位置に対応した収録位置で収録された音声データが含まれているものと判断し、音声符号化階層を決定してもよい。

【0062】

以上説明したように、本実施形態の再生装置は、階層符号化映像データの所定の符号化階層に対応付けられた音声データを、再生する符号化階層の復号映像の画角に合わせて補正することで、ユーザに臨場感を与える音声データの再生を可能とする。具体的には再生装置は、階層符号化映像データと、当該階層符号化映像データの所定の符号化階層に対応付けられた音声データとを受信する。再生装置はさらに階層符号化映像データの複数の符号化階層のうち、音声データが対応付けられている音声対応符号化階層を特定する。そして、音声対応符号化階層の復号映像の画角に対する、再生する符号化階層の復号映像の画角の比を算出し、予め記憶された音声補正量にさらに乗ずることにより、得られた新たな音声補正量を用いて音声データを補正する。

【0063】

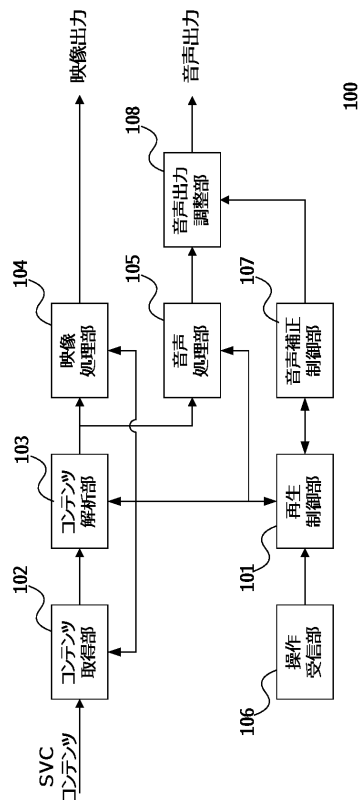
このようにすることで、選択された符号化階層に対応した音声が存在しない場合であっても、当該選択された符号化階層に対応した臨場感のある音声をユーザに提示可能である。

【0064】

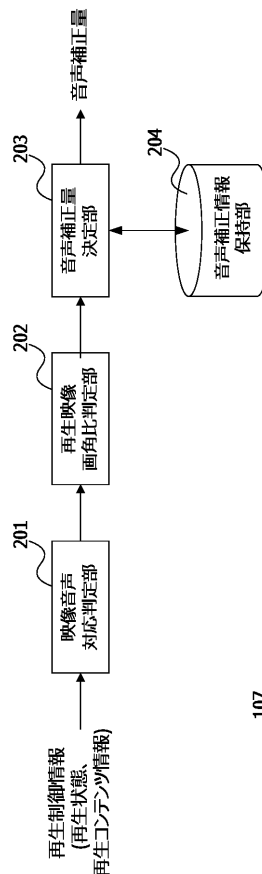
(その他の実施形態)

また、本発明は、以下の処理を実行することによっても実現される。即ち、上述した実施形態の機能を実現するソフトウェア(プログラム)を、ネットワーク又は各種記憶媒体を介してシステム或いは装置に供給し、そのシステム或いは装置のコンピュータ(またはCPUやMPU等)がプログラムを読み出して実行する処理である。

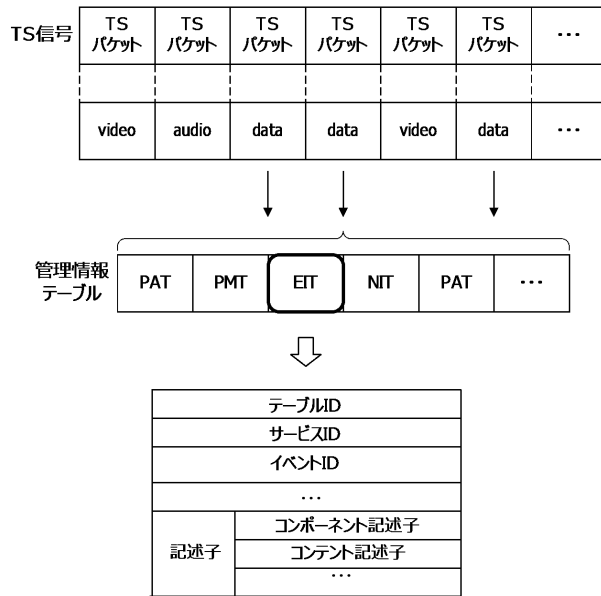
【図1】



【図2】



【図 3】



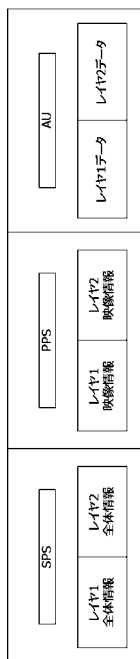
【図 4】

```

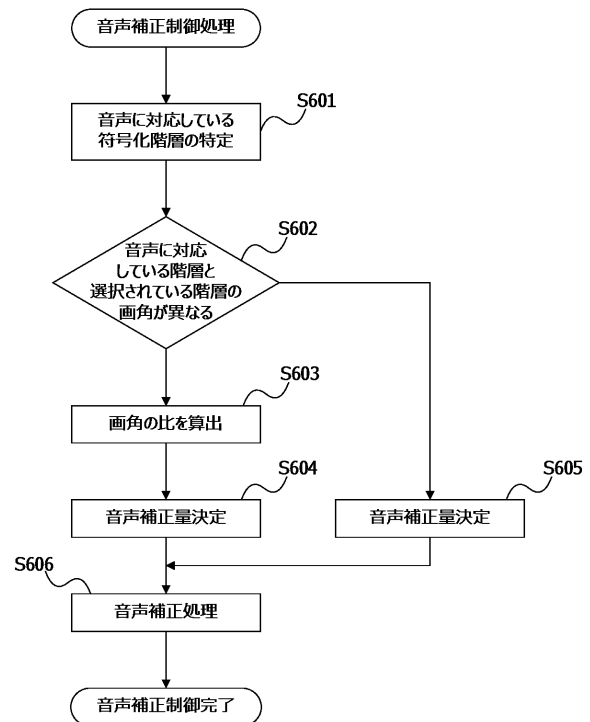
component_descriptor(){
  descriptor_tag
  descriptor_length
  reserved_future_use
  stream_content
  component_type
  default_layer
  component_tag
  ISO_639_language_code
  for (i=0;i<N;i++){
    text_char
  }
}

```

【図 5】



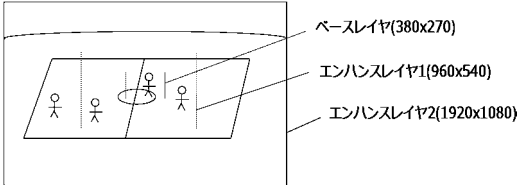
【図 6】



【図 7】

ジャンル	高音	低音	サラウンド
スポーツ	20	-20	20
ニュース	10	-10	10
音楽	10	-10	20

【図 8】



階層符号化映像
(画角:ベース<エンハンス1<エンハンス2)
(解像度:ベース<エンハンス1<エンハンス2)

フロントページの続き

- (72)発明者 占部 弘文
東京都大田区下丸子3丁目30番2号 キヤノン株式会社内
- (72)発明者 水留 敦
東京都大田区下丸子3丁目30番2号 キヤノン株式会社内

審査官 矢野 光治

- (56)参考文献 特開2006-287544(JP,A)
特開2009-290781(JP,A)
特開2007-018646(JP,A)
特開2005-318225(JP,A)
特開2004-032000(JP,A)

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