

明 細 書

発明の名称： 情報処理装置および情報処理方法

技術分野

[0001] 本開示は、情報処理装置および情報処理方法に関し、特に、複数の種類の音声データのうちの所定の種類の音声データの取得効率を向上させることができるようにした情報処理装置および情報処理方法に関する。

背景技術

[0002] 近年、インターネット上のストリーミングサービスの主流がOTT-V (Over The Top Video) となっている。この基盤技術として普及し始めているのがMP EG-DASH (Moving Picture Experts Group phase - Dynamic Adaptive Streaming over HTTP) である (例えば、非特許文献1 参照)。

[0003] MPEG-DASHでは、配信サーバが1本の動画コンテンツ用に画面サイズと符号化速度が異なる動画データ群を用意し、再生端末が伝送路の状況に応じて最適な画面サイズと符号化速度の動画データ群を要求することにより、適応型のストリーミング配信が実現される。

先行技術文献

非特許文献

[0004] 非特許文献1：MPEG-DASH(Dynamic Adaptive Streaming over HTTP) (URL:<http://mpeg.chiariglione.org/standards/mpeg-dash/media-presentation-description-and-segment-formats/text-isoiec-23009-12012-dam-1>)

発明の概要

発明が解決しようとする課題

[0005] しかしながら、動画コンテンツの複数の種類の音声データのうちの所定の種類の音声データの取得効率を向上させることは考えられていない。

[0006] 本開示は、このような状況に鑑みてなされたものであり、複数の種類の音声データのうちの所定の種類の音声データの取得効率を向上させることがで

きるようにするものである。

課題を解決するための手段

- [0007] 本開示の第1の側面の情報処理装置は、複数の種類の音声データが前記種類に応じて複数のトラックに分割されて配置されるファイルのうちの所定の前記トラックの音声データを取得する取得部を備える情報処理装置である。
- [0008] 本開示の第1の側面の情報処理方法は、本開示の第1の側面の情報処理装置に対応する。
- [0009] 本開示の第1の側面においては、複数の種類の音声データが前記種類に応じて複数のトラックに分割されて配置されるファイルのうちの所定の前記トラックの音声データが取得される。
- [0010] 本開示の第2の側面の情報処理装置は、複数の種類の音声データが前記種類に応じて複数のトラックに分割されて配置されるファイルを生成する生成部を備える情報処理装置である。
- [0011] 本開示の第2の側面の情報処理方法は、本開示の第2の側面の情報処理装置に対応する。
- [0012] 本開示の第2の側面においては、複数の種類の音声データが前記種類に応じて複数のトラックに分割されて配置されるファイルが生成される。
- [0013] なお、第1および第2の側面の情報処理装置は、コンピュータにプログラムを実行させることにより実現することができる。
- [0014] また、第1および第2の側面の情報処理装置を実現するために、コンピュータに実行させるプログラムは、伝送媒体を介して伝送することにより、又は、記録媒体に記録して、提供することができる。

発明の効果

- [0015] 本開示の第1の側面によれば、音声データを取得することができる。また、本開示の第1の側面によれば、複数の種類の音声データのうちの所定の種類の音声データを効率良く取得することができる。
- [0016] 本開示の第2の側面によれば、ファイルを生成することができる。また、本開示の第2の側面によれば、複数の種類の音声データのうちの所定の種類

の音声データの取得効率を向上させるファイルを生成することができる。

図面の簡単な説明

[0017] [図1]本開示を適用した情報処理システムの第1の例の概要を説明する図である。

[図2]タイルの例を示す図である。

[図3]オブジェクトを説明する図である。

[図4]オブジェクト位置情報を説明する図である。

[図5]画枠サイズ情報を説明する図である。

[図6]MPDファイルの構造を示す図である。

[図7]「Period」、「Representation」、および「Segment」の関係を示す図である。

[図8]MPDファイルの階層構造を示す図である。

[図9]MPDファイルの構造と時間軸との関係を示す図である。

[図10]MPDファイルの記述例を説明する図である。

[図11]ファイル生成装置の構成例を示すブロック図である。

[図12]ファイル生成装置のファイル生成処理を説明するフローチャートである。

[図13]ストリーミング再生部の構成例を示すブロック図である。

[図14]ストリーミング再生部のストリーミング再生処理を説明するフローチャートである。

[図15]MPDファイルの他の記述例を説明する図である。

[図16]MPDファイルのさらに他の記述例を説明する図である。

[図17]オーディオストリームの配置例を示す図である。

[図18]gsixの記述例を示す図である。

[図19]Sample group entryとオブジェクトIDの対応関係を示す情報の例を示す図である。

[図20]AudioObjectSampleGroupEntryの記述例を示す図である。

[図21]type assignment boxの記述例を示す図である。

[図22]本開示を適用した情報処理システムの第2の例の概要を説明する図である。

[図23]本開示を適用した情報処理システムのストリーミング再生部の構成例を示すブロック図である。

[図24]オブジェクトの位置の決定方法を説明する図である。

[図25]オブジェクトの位置の決定方法を説明する図である。

[図26]オブジェクトの位置の決定方法を説明する図である。

[図27]水平角度 θ_{Ai} と水平角度 θ_{Ai}' の関係を表す図である。

[図28]図23のストリーミング再生部のストリーミング再生処理を説明するフローチャートである。

[図29]図28の位置決定処理の詳細を説明するフローチャートである。

[図30]図29の水平角度 θ_{Ai}' 推定処理の詳細を説明するフローチャートである。

[図31]MP4の3Dオーディオファイルフォーマットのトラックの概要を説明する図である。

[図32]moovボックスの構造を示す図である。

[図33]本開示を適用した第1実施の形態におけるトラックの概要を説明する図である。

[図34]図33のベーストラックのサンプルエントリのシンタクスの例を示す図である。

[図35]図33のチャンネルオーディオトラックのサンプルエントリのシンタクスの例を示す図である。

[図36]図33のオブジェクトオーディオトラックのサンプルエントリのシンタクスの例を示す図である。

[図37]図33のH0Aオーディオトラックのサンプルエントリのシンタクスの例を示す図である。

[図38]図33のオブジェクトメタデータトラックのサンプルエントリのシンタクスの例を示す図である。

[図39]セグメント構造の第1の例を示す図である。

[図40]セグメント構造の第2の例を示す図である。

[図41]level assignmentボックスの記述例を示す図である。

[図42]本開示を適用した第1実施の形態におけるMPDファイルの記述例を示す図である。

[図43]essential Propertyの定義を示す図である。

[図44]本開示を適用した第1実施の形態における情報処理システムの概要を説明する図である。

[図45]図44のファイル生成装置の構成例を示すブロック図である。

[図46]図45のファイル生成装置のファイル生成処理を説明するフローチャートである。

[図47]図44の動画再生端末により実現されるストリーミング再生部の構成例を示すブロック図である。

[図48]図47のストリーミング再生部のチャンネルオーディオ再生処理を説明するフローチャートである。

[図49]図47のストリーミング再生部のオブジェクト特定処理を説明するフローチャートである。

[図50]図47のストリーミング再生部の特定オブジェクトオーディオ再生処理を説明するフローチャートである。

[図51]本開示を適用した第2実施の形態におけるトラックの概要を説明する図である。

[図52]図51のベーストラックのサンプルエントリのシンタクスの例を示す図である。

[図53]ベースサンプルの構造を示す図である。

[図54]ベースのサンプルのシンタクスの例を示す図である。

[図55]extractorのデータの例を示す図である。

[図56]本開示を適用した第3実施の形態におけるトラックの概要を説明する図である。

[図57]本開示を適用した第4実施の形態におけるトラックの概要を説明する図である。

[図58]本開示を適用した第4実施の形態におけるMPDファイルの記述例を示す図である。

[図59]本開示を適用した第4実施の形態における情報処理システムの概要を説明する図である。

[図60]図59のファイル生成装置の構成例を示すブロック図である。

[図61]図60のファイル生成装置のファイル生成処理を説明するフローチャートである。

[図62]図59の動画再生端末により実現されるストリーミング再生部の構成例を示すブロック図である。

[図63]図62のストリーミング再生部のチャンネルオーディオ再生処理の例を説明するフローチャートである。

[図64]図62のストリーミング再生部のオブジェクトオーディオ再生処理の第1の例を説明するフローチャートである。

[図65]図62のストリーミング再生部のオブジェクトオーディオ再生処理の第2の例を説明するフローチャートである。

[図66]図62のストリーミング再生部のオブジェクトオーディオ再生処理の第3の例を説明するフローチャートである。

[図67]プライオリティに基づいて選択されたオブジェクトの例を示す図である。

[図68]本開示を適用した第5実施の形態におけるトラックの概要を説明する図である。

[図69]本開示を適用した第6実施の形態におけるトラックの概要を説明する図である。

[図70]3Dオーディオの階層構造を示す図である。

[図71]Webサーバの処理の第1の例を説明する図である。

[図72]Webサーバのトラック分割処理を説明するフローチャートである。

[図73]音声復号処理部の処理の第1の例を説明する図である。

[図74]音声復号処理部の復号処理の第1の例の詳細を説明するフローチャートである。

[図75]音声復号処理部の処理の第2の例を説明する図である。

[図76]音声復号処理部の復号処理の第2の例の詳細を説明するフローチャートである。

[図77]Webサーバの処理の第2の例を説明する図である。

[図78]音声復号処理部の処理の第3の例を説明する図である。

[図79]音声復号処理部の復号処理の第3の例の詳細を説明するフローチャートである。

[図80]ベースサンプルに配置されるConfig情報のシンタクスの第2の例を示す図である。

[図81]図80のExtエレメント用のConfig情報のシンタクスの例を示す図である。

[図82]図81のExtractor用のConfig情報のシンタクスの例を示す図である。

[図83]ベースサンプルに配置されるフレーム単位のデータのシンタクスの第2の例を示す図である。

[図84]図83のExtractorのデータのシンタクスの例を示す図である。

[図85]ベースサンプルに配置されるConfig情報のシンタクスの第3の例を示す図である。

[図86]ベースサンプルに配置されるフレーム単位のデータのシンタクスの第3の例を示す図である。

[図87]本開示を適用した情報処理システムの第7実施の形態におけるオーディオストリームの構成例を示す図である。

[図88]第7実施の形態におけるトラックの概要を説明する図である。

[図89]第7実施の形態におけるファイル生成処理を説明するフローチャートである。

[図90]第7実施の形態におけるオーディオ再生処理を説明するフローチャートである。

トである。

[図91]本開示を適用した情報処理システムの第8実施の形態におけるトラックの概要を説明する図である。

[図92]音声ファイルの構成例を示す図である。

[図93]音声ファイルの他の構成例を示す図である。

[図94]音声ファイルのさらに他の構成例を示す図である。

[図95]コンピュータのハードウェアの構成例を示すブロック図である。

発明を実施するための形態

[0018] 以下、本開示の前提および本開示を実施するための形態（以下、実施の形態という）について説明する。なお、説明は以下の順序で行う。

0. 本開示の前提（図1乃至図30）
1. 第1実施の形態（図31乃至図50）
2. 第2実施の形態（図51乃至図55）
3. 第3実施の形態（図56）
4. 第4実施の形態（図57乃至図67）
5. 第5実施の形態（図68）
6. 第6実施の形態（図69）
7. 3Dオーディオの階層構造の説明（図70）
8. Webサーバの処理の第1の例の説明（図71および図72）
9. 音声復号処理部の処理の第1の例の説明（図73および図74）
10. 音声復号処理部の処理の第2の例の説明（図75および図76）
11. Webサーバの処理の第2の例の説明（図77）
12. 音声復号処理部の処理の第3の例の説明（図78および図79）
13. ベースサンプルのシンタクスの第2の例（図80乃至図84）
14. ベースサンプルのシンタクスの第3の例（図85および図86）
15. 第7実施の形態（図87乃至図90）
16. 第8実施の形態（図91乃至図94）
17. 第9実施の形態（図95）

[0019] <本開示の前提>

(情報処理システムの第1の例の概要)

図1は、本開示を適用した情報処理システムの第1の例の概要を説明する図である。

[0020] 図1の情報処理システム10は、ファイル生成装置11と接続されるWebサーバ12と動画再生端末14が、インターネット13を介して接続されることにより構成される。

[0021] 情報処理システム10では、MPEG-DASHに準ずる方式で、Webサーバ12が動画コンテンツの画像データをタイル単位で動画再生端末14に配信（タイルドストリーミング）する。

[0022] 具体的には、ファイル生成装置11は、動画コンテンツの画像データを取得し、タイル単位で符号化してビデオストリームを生成する。ファイル生成装置11は、各タイルのビデオストリームを、セグメントと呼ばれる数秒から10秒程度の時間単位ごとにファイル化する。ファイル生成装置11は、その結果得られる各タイルの画像ファイルをWebサーバ12にアップロードする。

[0023] また、ファイル生成装置11は、動画コンテンツの音声データをオブジェクト（詳細は後述する）ごとに取得し、オブジェクト単位で符号化してオーディオストリームを生成する。ファイル生成装置11は、各オブジェクトのオーディオストリームをセグメント単位でファイル化し、その結果得られる各オブジェクトの音声ファイルをWebサーバ12にアップロードする。

[0024] なお、オブジェクトとは、音源であり、各オブジェクトの音声データは、そのオブジェクトに取り付けられたマイクロフォンなどにより取得される。オブジェクトは、固定されたマイクスタンドなどの物体であってもよいし、人物などの動体であってもよい。

[0025] ファイル生成装置11は、各オブジェクトの位置（音声データの取得位置）を表すオブジェクト位置情報（音声位置情報）、オブジェクトに固有のIDであるオブジェクトID等を含むオーディオメタデータを符号化する。フ

ファイル生成装置 11 は、オーディオメタデータの符号化データをセグメント単位でファイル化し、その結果得られる音声メタファイルを Webサーバ 12 にアップロードする。

[0026] また、ファイル生成装置 11 は、動画コンテンツの画像の画枠サイズを表す画枠サイズ情報、各タイルの画像上の位置を表すタイル位置情報等を含む、画像ファイルや音声ファイルを管理する MPD (Media Presentation Description) ファイル (制御情報) を生成する。ファイル生成装置 11 は、MPD ファイルを Webサーバ 12 にアップロードする。

[0027] Webサーバ 12 は、ファイル生成装置 11 からアップロードされた画像ファイル、音声ファイル、音声メタファイル、および MPD ファイルを格納する。

[0028] 図 1 の例では、Webサーバ 12 には、タイル ID 「1」のタイルの複数のセグメントの画像ファイルからなるセグメント群と、タイル ID 「2」のタイルの複数のセグメントの画像ファイルからなるセグメント群とが格納されている。また、Webサーバ 12 には、オブジェクト ID 「1」のオブジェクトの複数のセグメントの音声ファイルからなるセグメント群と、オブジェクト ID 「2」のオブジェクトの複数のセグメントの音声ファイルからなるセグメント群とが格納されている。図示は省略するが、音声メタファイルのセグメント群も同様に格納されている。

[0029] なお、以下では、タイル ID が i であるタイルをタイル # i といい、オブジェクト ID が i のオブジェクトをオブジェクト # i という。

[0030] Webサーバ 12 は、伝送部として機能し、動画再生端末 14 からの要求に応じて、格納している画像ファイル、音声ファイル、音声メタファイル、MPD ファイル等を動画再生端末 14 に送信する。

[0031] 動画再生端末 14 は、ストリーミングデータの制御用ソフトウェア (以下、制御用ソフトウェアという) 21、動画再生ソフトウェア 22、HTTP (HyperText Transfer Protocol) アクセス用のクライアント・ソフトウェア (以下、アクセス用ソフトウェアという) 23 などを実行する。

[0032] 制御用ソフトウェア 21 は、Webサーバ 12 からストリーミングするデータ

を制御するソフトウェアである。具体的には、制御用ソフトウェア 21 は、動画再生端末 14 に Web サーバ 12 から MPD ファイルを取得させる。

[0033] また、制御用ソフトウェア 21 は、動画再生ソフトウェア 22 から指示される動画コンテンツの画像内の表示する領域である表示領域と、MPD ファイルに含まれるタイル位置情報とに基づいて、表示領域内のタイルを特定する。そして、制御用ソフトウェア 21 は、そのタイルの画像ファイルの送信要求をアクセス用ソフトウェア 23 に指令する。

[0034] また、制御用ソフトウェア 21 は、アクセス用ソフトウェア 23 に音声メタファイルの送信要求を指令する。そして、制御用ソフトウェア 21 は、表示領域、MPD ファイルに含まれる画枠サイズ情報、および音声メタファイルに含まれるオブジェクト位置情報に基づいて、表示領域内の画像に対応するオブジェクトを特定する。そして、制御用ソフトウェア 21 は、そのオブジェクトの音声ファイルの送信要求をアクセス用ソフトウェア 23 に指令する。

[0035] 動画再生ソフトウェア 22 は、Web サーバ 12 から取得された画像ファイルと音声ファイルを再生するソフトウェアである。具体的には、動画再生ソフトウェア 22 は、ユーザにより表示領域が指定されると、その表示領域を制御用ソフトウェア 21 に指示する。また、動画再生ソフトウェア 22 は、その指示に応じて Web サーバ 12 から取得された画像ファイルと音声ファイルを復号し、合成して出力する。

[0036] アクセス用ソフトウェア 23 は、HTTP を用いたインターネット 13 を介した Web サーバ 12 との通信を制御するソフトウェアである。具体的には、アクセス用ソフトウェア 23 は、制御用ソフトウェア 21 の指令に応じて、画像ファイル、音声ファイル、および音声メタファイルの送信要求を動画再生端末 14 に送信させる。また、アクセス用ソフトウェア 23 は、その送信要求に応じて、Web サーバ 12 から送信されてくる画像ファイル、音声ファイル、および音声メタファイルを動画再生端末 14 に受信させる。

[0037] (タイルの例)

図 2 は、タイルの例を示す図である。

[0038] 図2に示すように、動画コンテンツの画像は、複数のタイルに分割され、各タイルに1から順にタイルIDが付与される。図2の例では、動画コンテンツの画像は、4つのタイル#1乃至タイル#4に分割されている。

[0039] (オブジェクトの説明)

図3は、オブジェクトを説明する図である。

[0040] 図3の例では、動画コンテンツの音声として、画像内の8つのオブジェクトの音声取得されており、各オブジェクトには1から順にオブジェクトIDが付与されている。オブジェクト#1乃至オブジェクト#5は、それぞれ、動体であり、オブジェクト#6乃至オブジェクト#8は、固定された物体である。また、図3の例では、動画コンテンツの画像は、5（縦）×7（横）のタイルに分割されている。

[0041] この場合、図3に示すように、ユーザにより2（縦）×3（横）個のタイルからなる表示領域31が指定されると、表示領域31には、オブジェクト#1、オブジェクト#2、およびオブジェクト#6のみが含まれる。従って、動画再生端末14は、例えば、オブジェクト#1、オブジェクト#2、およびオブジェクト#6の音声ファイルのみをWebサーバ12から取得し、再生する。

[0042] 表示領域31内のオブジェクトは、以下に説明するように、画枠サイズ情報およびオブジェクト位置情報に基づいて特定することができる。

[0043] (オブジェクト位置情報の説明)

図4は、オブジェクト位置情報を説明する図である。

[0044] 図4に示すように、オブジェクト位置情報は、オブジェクト40の水平角度 θ_A ($-180^\circ \leq \theta_A \leq 180^\circ$)、垂直角度 γ_A ($-90^\circ \leq \gamma_A \leq 90^\circ$)、および距離 r_A ($0 < r_A$) からなる。水平角度 θ_A は、例えば、画像の中心の撮影位置を原点（基点）Oとし、画像の水平方向をX方向、垂直方向をY方向、XY平面に垂直な奥行き方向をZ方向としたときの、オブジェクト40と原点Oを結ぶ直線とYZ平面との水平方向の角度である。垂直角度 γ_A は、オブジェクト40と原点Oを結ぶ直線とXZ平面との垂直方向の角度であり、距離 r_A は、オブジ

エクト40と原点Oの距離である。

[0045] なお、本明細書では、左回転および上回転の角度を正の角度とし、右回転および下回転の角度を負の角度とする。

[0046] (画枠サイズ情報の説明)

図5は、画枠サイズ情報を説明する図である。

[0047] 図5に示すように、画枠サイズ情報は、画枠の左端の水平角度 θ_{v1} 、右端の水平角度 θ_{v2} 、画枠の上端の垂直角度 γ_{v1} 、下端の垂直角度 γ_{v2} 、および距離 r_v により構成される。

[0048] 水平角度 θ_{v1} は、例えば、画像の中心の撮影位置を原点Oとし、画像の水平方向をX方向、垂直方向をY方向、XY平面に垂直な奥行き方向をZ方向としたときの、画枠の左端と原点Oを結ぶ直線とYZ平面との水平方向の角度である。水平角度 θ_{v2} は、画枠の右端と原点Oを結ぶ直線とYZ平面との水平方向の角度である。従って、水平角度 θ_{v1} と水平角度 θ_{v2} とを合わせた角度は、水平方向の画角である。

[0049] 垂直角度 γ_{v1} 、 γ_{v2} は、それぞれ、画枠の上端、下端と原点Oを結ぶ直線とXZ平面との角度であり、垂直角度 γ_{v1} と垂直角度 γ_{v2} とを合わせた角度は、垂直方向の画角である。距離 r_v は、原点Oと画像の平面との距離である。

[0050] 以上のように、オブジェクト位置情報と画枠サイズ情報は、それぞれ、オブジェクト40、画枠と、原点Oとの位置関係を表している。従って、オブジェクト位置情報と画枠サイズ情報に基づいて、各オブジェクトの画像上の位置を検出(認識)することができる。その結果、表示領域31内のオブジェクトを特定することができる。

[0051] (MPDファイルの構造の説明)

図6は、MPDファイルの構造を示す図である。

[0052] MPDファイルの解析(パース)においては、動画再生端末14は、MPDファイル(図6のMedia Presentation)の「Period」に含まれる「Representation」の属性から最適なものを選択する。

[0053] 動画再生端末14は、選択した「Representation」の先頭の「Initializat

ion Segment」のURL (Uniform Resource Locator) 等を参照してファイルを取得し、処理する。続いて、動画再生端末 14 は、後続の「Media Segment」のURL等を参照してファイルを取得し、再生する。

[0054] なお、MPDファイルにおける、「Period」、「Representation」、および「Segment」の関係は、図7のようになる。つまり、1つの動画コンテンツは、「Period」により、セグメントより長い時間単位で管理することができ、各「Period」において、「Segment」によりセグメント単位で管理することができる。また、各「Period」において、「Representation」により、動画コンテンツをストリームの属性単位で管理することができる。

[0055] したがって、MPDファイルは、「Period」以下において、図8に示す階層構造を有する。また、このMPDファイルの構造を時間軸上に並べると図9の例のようになる。図9から明らかなように、同一のセグメントに対して複数の「Representation」が存在している。動画再生端末 14 は、これらのうちのいずれかを適応的に選択することにより、ユーザにより選択された表示領域の画像ファイルと音声ファイルを取得し、再生することができる。

[0056] (MPDファイルの記述の説明)

図10は、MPDファイルの記述を説明する図である。

[0057] 上述したように、情報処理システム 10 では、動画再生端末 14 において表示領域内のオブジェクトを特定可能にするため、MPDファイルに画枠サイズ情報が含まれる。この画枠サイズ情報は、図10に示すように、ViewpointのDescriptorType elementを活用して新たに画枠サイズ情報 (Viewing Angle) を定義するためのScheme(urn:mpeg:DASH:viewingAngle:2013)を拡張することにより、音声用の「Adaptation Set」と画像用の「Adaptation Set」に配置される。画枠サイズ情報は、画像用の「Adaptation Set」にのみ配置されるようにしてもよい。

[0058] また、MPDファイルの音声用の「Adaptation Set」には、音声メタファイル用の「Representation」が記述され、その「Representation」の「Segment」には、音声メタファイル (audionmetadata.mp4) を特定する情報としてのURL

等が記述される。このとき、Role elementを活用して、「Segment」で指定されるファイルが音声メタファイルであること（objectaudiometadata）が記述される。

[0059] MPDファイルの音声用の「Adaptation Set」にはまた、各オブジェクトの音声ファイル用の「Representation」が記述され、その「Representation」の「Segment」には、各オブジェクトの音声ファイル（audio0bje1.mp4, audio0bje5.mp4）を特定する情報としてのURL等が記述される。このとき、Viewpointを拡張することにより、音声ファイルに対応するオブジェクトのオブジェクトID（1,5）も記述される。

[0060] なお、図示は省略するが、タイル位置情報は、画像用の「Adaptation Set」に配置される。

[0061] （ファイル生成装置の構成例）

図11は、図1のファイル生成装置11の構成例を示すブロック図である。

[0062] 図11のファイル生成装置11は、画面分割処理部51、画像符号化処理部52、画像ファイル生成部53、画像情報生成部54、音声符号化処理部55、音声ファイル生成部56、MPD生成部57、およびサーバアップロード処理部58により構成される。

[0063] ファイル生成装置11の画面分割処理部51は、外部から入力される動画コンテンツの画像データをタイル単位に分割する。画面分割処理部51は、タイル位置情報を画像情報生成部54に供給する。また、画面分割処理部51は、タイル単位の画像データを画像符号化処理部52に供給する。

[0064] 画像符号化処理部52は、画面分割処理部51から供給されるタイル単位の画像データをタイルごとに符号化し、ビデオストリームを生成する。画像符号化処理部52は、各タイルのビデオストリームを画像ファイル生成部53に供給する。

[0065] 画像ファイル生成部53は、画像符号化処理部52から供給される各タイルのビデオストリームをセグメント単位でファイル化し、その結果得られる

各タイルの画像ファイルをMPD生成部57に供給する。

- [0066] 画像情報生成部54は、画面分割処理部51から供給されるタイル位置情報と外部から入力される画枠サイズ情報とを画像情報としてMPD生成部57に供給する。
- [0067] 音声符号化処理部55は、外部から入力される動画コンテンツのオブジェクト単位の音声データをオブジェクトごとに符号化し、オーディオストリームを生成する。また、音声符号化処理部55は、外部から入力される各オブジェクトのオブジェクト位置情報、オブジェクトID等を含むオーディオメタデータを符号化し、符号化データを生成する。音声符号化処理部55は、各オブジェクトのオーディオストリームとオーディオメタデータの符号化データとを音声ファイル生成部56に供給する。
- [0068] 音声ファイル生成部56は、音声ファイル生成部として機能し、音声符号化処理部55から供給される各オブジェクトのオーディオストリームをセグメント単位でファイル化し、その結果得られる各オブジェクトの音声ファイルをMPD生成部57に供給する。
- [0069] また、音声ファイル生成部56は、メタファイル生成部として機能し、音声符号化処理部55から供給されるオーディオメタデータの符号化データをセグメント単位でファイル化し、その結果得られる音声メタファイルをMPD生成部57に供給する。
- [0070] MPD生成部57は、画像ファイル生成部53から供給される各タイルの画像ファイルを格納するWebサーバ12のURL等を決定する。また、MPD生成部57は、音声ファイル生成部56から供給される各オブジェクトの音声ファイルと音声メタファイルを格納するWebサーバ12のURL等を決定する。
- [0071] MPD生成部57は、画像情報生成部54から供給される画像情報をMPDファイルの画像用の「AdaptationSet」に配置する。また、MPD生成部57は、画像情報のうちの画枠サイズ情報をMPDファイルの音声用の「AdaptationSet」に配置する。MPD生成部57は、各タイルの画像ファイルのURL等を、そのタイルの画像ファイル用の「Representation」の「Segment」に配置する。

- [0072] MPD生成部57は、各オブジェクトの音声ファイルのURL等を、そのオブジェクトの音声ファイル用の「Representation」の「Segment」に配置する。また、MPD生成部57は、情報生成部として機能し、音声メタファイルを特定する情報としてのURL等を音声メタファイル用の「Representation」の「Segment」に配置する。MPD生成部57は、以上のようにして各種の情報が配置されたMPDファイル、並びに、画像ファイル、音声ファイル、および音声メタファイルをサーバアップロード処理部58に供給する。
- [0073] サーバアップロード処理部58は、MPD生成部57から供給される各タイルの画像ファイル、各オブジェクトの音声ファイル、音声メタファイル、およびMPDファイルを、Webサーバ12にアップロードする。
- [0074] (ファイル生成装置の処理の説明)
- 図12は、図11のファイル生成装置11のファイル生成処理を説明するフローチャートである。
- [0075] 図12のステップS11において、ファイル生成装置11の画面分割処理部51は、外部から入力される動画コンテンツの画像データをタイル単位に分割する。画面分割処理部51は、タイル位置情報を画像情報生成部54に供給する。また、画面分割処理部51は、タイル単位の画像データを画像符号化処理部52に供給する。
- [0076] ステップS12において、画像符号化処理部52は、画面分割処理部51から供給されるタイル単位の画像データをタイルごとに符号化し、各タイルのビデオストリームを生成する。画像符号化処理部52は、各タイルのビデオストリームを画像ファイル生成部53に供給する。
- [0077] ステップS13において、画像ファイル生成部53は、画像符号化処理部52から供給される各タイルのビデオストリームをセグメント単位でファイル化し、各タイルの画像ファイルを生成する。画像ファイル生成部53は、各タイルの画像ファイルをMPD生成部57に供給する。
- [0078] ステップS14において、画像情報生成部54は、外部から画枠サイズ情報を取得する。ステップS15において、画像情報生成部54は、画面分割

処理部 5 1 から供給されるタイル位置情報と画枠サイズ情報とを含む画像情報を生成し、MPD生成部 5 7 に供給する。

[0079] ステップ S 1 6 において、音声符号化処理部 5 5 は、外部から入力される動画コンテンツのオブジェクト単位の音声データをオブジェクトごとに符号化し、各オブジェクトのオーディオストリームを生成する。また、音声符号化処理部 5 5 は、外部から入力される各オブジェクトのオブジェクト位置情報、オブジェクト ID 等を含むオーディオメタデータを符号化し、符号化データを生成する。音声符号化処理部 5 5 は、各オブジェクトのオーディオストリームとオーディオメタデータの符号化データとを音声ファイル生成部 5 6 に供給する。

[0080] ステップ S 1 7 において、音声ファイル生成部 5 6 は、音声符号化処理部 5 5 から供給される各オブジェクトのオーディオストリームをセグメント単位でファイル化し、各オブジェクトの音声ファイルを生成する。また、音声ファイル生成部 5 6 は、音声符号化処理部 5 5 から供給されるオーディオメタデータの符号化データをセグメント単位でファイル化し、音声メタファイルを生成する。音声ファイル生成部 5 6 は、各オブジェクトの音声ファイルと音声メタファイルをMPD生成部 5 7 に供給する。

[0081] ステップ S 1 8 において、MPD生成部 5 7 は、画像情報生成部 5 4 から供給される画像情報、各ファイルのURL等を含むMPDファイルを生成する。MPD生成部 5 7 は、MPDファイル、各タイルの画像ファイル、各オブジェクトの音声ファイル、および音声メタファイルをサーバアップロード処理部 5 8 に供給する。

[0082] ステップ S 1 9 において、サーバアップロード処理部 5 8 は、MPD生成部 5 7 から供給される各タイルの画像ファイル、各オブジェクトの音声ファイル、音声メタファイル、およびMPDファイルを、Webサーバ 1 2 にアップロードする。そして、処理は終了する。

[0083] (動画再生端末の機能的構成例)

図 1 3 は、図 1 の動画再生端末 1 4 が制御用ソフトウェア 2 1、動画再生

ソフトウェア 22、およびアクセス用ソフトウェア 23 を実行することにより実現されるストリーミング再生部の構成例を示すブロック図である。

[0084] 図 13 のストリーミング再生部 90 は、MPD 取得部 91、MPD 処理部 92、メタファイル取得部 93、音声選択部 94、音声ファイル取得部 95、音声復号処理部 96、音声合成処理部 97、画像選択部 98、画像ファイル取得部 99、画像復号処理部 100、および画像合成処理部 101 により構成される。

[0085] ストリーミング再生部 90 の MPD 取得部 91 は、受け取り部として機能し、Web サーバ 12 から MPD ファイルを取得し、MPD 処理部 92 に供給する。

[0086] MPD 処理部 92 は、MPD 取得部 91 から供給される MPD ファイルから音声メタファイル用の「Segment」に記述される URL 等の情報を抽出し、メタファイル取得部 93 に供給する。また、MPD 処理部 92 は、MPD ファイルから画像用の「AdaptationSet」に記述される画枠サイズ情報を抽出し、音声選択部 94 に供給する。MPD 処理部 92 は、MPD ファイルから、音声選択部 94 から要求されるオブジェクトの音声ファイル用の「Segment」に記述される URL 等の情報を抽出し、音声選択部 94 に供給する。

[0087] MPD 処理部 92 は、MPD ファイルから画像用の「AdaptationSet」に記述されるタイル位置情報を抽出し、画像選択部 98 に供給する。MPD 処理部 92 は、MPD ファイルから、画像選択部 98 から要求されるタイルの画像ファイル用の「Segment」に記述される URL 等の情報を抽出し、画像選択部 98 に供給する。

[0088] メタファイル取得部 93 は、MPD 処理部 92 から供給される URL 等の情報に基づいて、その URL で特定される音声メタファイルを Web サーバ 12 に要求し、取得する。メタファイル取得部 93 は、音声メタファイルに含まれるオブジェクト位置情報を音声選択部 94 に供給する。

[0089] 音声選択部 94 は、位置決定部として機能し、MPD 処理部 92 から供給される画枠サイズ情報とメタファイル取得部 93 から供給されるオブジェクト位置情報とに基づいて、各オブジェクトの画像上の位置を算出する。音声選択

部 9 4 は、各オブジェクトの画像上の位置に基づいて、ユーザにより指定される表示領域内のオブジェクトを選択する。音声選択部 9 4 は、選択されたオブジェクトの音声ファイルの URL 等の情報を MPD 処理部 9 2 に要求する。音声選択部 9 4 は、その要求に応じて MPD 処理部 9 2 から供給される URL 等の情報を音声ファイル取得部 9 5 に供給する。

[0090] 音声ファイル取得部 9 5 は、受け取り部として機能し、音声選択部 9 4 から供給される URL 等の情報に基づいて、その URL で特定されるオブジェクト単位の音声ファイルを Web サーバ 1 2 に要求し、取得する。音声ファイル取得部 9 5 は、取得されたオブジェクト単位の音声ファイルを音声復号処理部 9 6 に供給する。

[0091] 音声復号処理部 9 6 は、音声ファイル取得部 9 5 から供給されるオブジェクト単位の音声ファイルに含まれるオーディオストリームを復号し、オブジェクト単位の音声データを生成する。音声復号処理部 9 6 は、そのオブジェクト単位の音声データを音声合成処理部 9 7 に供給する。

[0092] 音声合成処理部 9 7 は、音声復号処理部 9 6 から供給されるオブジェクト単位の音声データを合成し、出力する。

[0093] 画像選択部 9 8 は、MPD 処理部 9 2 から供給されるタイル位置情報に基づいて、ユーザにより指定される表示領域内のタイルを選択する。画像選択部 9 8 は、選択されたタイルの画像ファイルの URL 等の情報を MPD 処理部 9 2 に要求する。画像選択部 9 8 は、その要求に応じて MPD 処理部 9 2 から供給される URL 等の情報を画像ファイル取得部 9 9 に供給する。

[0094] 画像ファイル取得部 9 9 は、画像選択部 9 8 から供給される URL 等の情報に基づいて、その URL で特定されるタイル単位の画像ファイルを Web サーバ 1 2 に要求し、取得する。画像ファイル取得部 9 9 は、取得されたタイル単位の画像ファイルを画像復号処理部 1 0 0 に供給する。

[0095] 画像復号処理部 1 0 0 は、画像ファイル取得部 9 9 から供給されるタイル単位の画像ファイルに含まれるビデオストリームを復号し、タイル単位の画像データを生成する。画像復号処理部 1 0 0 は、そのタイル単位の画像デー

タを画像合成処理部 101 に供給する。

[0096] 画像合成処理部 101 は、画像復号処理部 100 から供給されるタイル単位の画像データを合成し、出力する。

[0097] (動画再生端末の処理の説明)

図 14 は、動画再生端末 14 のストリーミング再生部 90 (図 13) のストリーミング再生処理を説明するフローチャートである。

[0098] 図 14 のステップ S31 において、ストリーミング再生部 90 の MPD 取得部 91 は、Web サーバ 12 から MPD ファイルを取得し、MPD 処理部 92 に供給する。

[0099] ステップ S32 において、MPD 処理部 92 は、MPD 取得部 91 から供給される MPD ファイルから、画像用の「AdaptationSet」に記述される画枠サイズ情報とタイル位置情報を取得する。MPD 処理部 92 は、画枠サイズ情報を音声選択部 94 に供給し、タイル位置情報を画像選択部 98 に供給する。また、MPD 処理部 92 は、音声メタファイル用の「Segment」に記述される URL 等の情報を抽出し、メタファイル取得部 93 に供給する。

[0100] ステップ S33 において、メタファイル取得部 93 は、MPD 処理部 92 から供給される URL 等の情報に基づいて、その URL で特定される音声メタファイルを Web サーバ 12 に要求し、取得する。メタファイル取得部 93 は、音声メタファイルに含まれるオブジェクト位置情報を音声選択部 94 に供給する。

[0101] ステップ S34 において、音声選択部 94 は、MPD 処理部 92 から供給される画枠サイズ情報とメタファイル取得部 93 から供給されるオブジェクト位置情報とに基づいて、ユーザにより指定される表示領域内のオブジェクトを選択する。音声選択部 94 は、選択されたオブジェクトの音声ファイルの URL 等の情報を MPD 処理部 92 に要求する。

[0102] MPD 処理部 92 は、MPD ファイルから、音声選択部 94 から要求されるオブジェクトの音声ファイル用の「Segment」に記述される URL 等の情報を抽出し、音声選択部 94 に供給する。音声選択部 94 は、MPD 処理部 92 から供給される URL 等の情報を音声ファイル取得部 95 に供給する。

- [0103] ステップS 3 5において、音声ファイル取得部 9 5は、音声選択部 9 4から供給されるURL等の情報に基づいて、そのURLで特定される、選択されたオブジェクトの音声ファイルをWebサーバ 1 2に要求し、取得する。音声ファイル取得部 9 5は、取得されたオブジェクト単位の音声ファイルを音声復号処理部 9 6に供給する。
- [0104] ステップS 3 6において、画像選択部 9 8は、MPD処理部 9 2から供給されるタイル位置情報に基づいて、ユーザにより指定される表示領域内のタイルを選択する。画像選択部 9 8は、選択されたタイルの画像ファイルのURL等の情報をMPD処理部 9 2に要求する。
- [0105] MPD処理部 9 2は、MPDファイルから、画像選択部 9 8から要求されるオブジェクトの画像ファイル用の「Segment」に記述されるURL等の情報を抽出し、画像選択部 9 8に供給する。画像選択部 9 8は、MPD処理部 9 2から供給されるURL等の情報を画像ファイル取得部 9 9に供給する。
- [0106] ステップS 3 7において、画像ファイル取得部 9 9は、画像選択部 9 8から供給されるURL等の情報に基づいて、そのURLで特定される、選択されたタイルの画像ファイルをWebサーバ 1 2に要求し、取得する。画像ファイル取得部 9 9は、取得されたタイル単位の画像ファイルを画像復号処理部 1 0 0に供給する。
- [0107] ステップS 3 8において、音声復号処理部 9 6は、音声ファイル取得部 9 5から供給されるオブジェクト単位の音声ファイルに含まれるオーディオストリームを復号し、オブジェクト単位の音声データを生成する。音声復号処理部 9 6は、そのオブジェクト単位の音声データを音声合成処理部 9 7に供給する。
- [0108] ステップS 3 9において、画像復号処理部 1 0 0は、画像ファイル取得部 9 9から供給されるタイル単位の画像ファイルに含まれるビデオストリームを復号し、タイル単位の画像データを生成する。画像復号処理部 1 0 0は、そのタイル単位の画像データを画像合成処理部 1 0 1に供給する。
- [0109] ステップS 4 0において、音声合成処理部 9 7は、音声復号処理部 9 6か

ら供給されるオブジェクト単位の音声データを合成し、出力する。ステップ S 4 1 において、画像合成処理部 1 0 1 は、画像復号処理部 1 0 0 から供給されるタイル単位の画像データを合成し、出力する。そして、処理は終了する。

[0110] 以上のように、Webサーバ 1 2 は画枠サイズ情報とオブジェクト位置情報を伝送する。従って、動画再生端末 1 4 は、例えば、表示領域内のオブジェクトを特定し、そのオブジェクトの音声ファイルを、表示領域内の画像に対応する音声ファイルとして選択的に取得することができる。このように、動画再生端末 1 4 は必要な音声ファイルのみを取得することができるので、伝送効率は向上する。

[0111] なお、図 1 5 に示すように、MPDファイルの画像用の「AdaptationSet」には、その画像と同時に再生したい音声に対応するオブジェクトを特定する情報としてのオブジェクト I D（オブジェクト特定情報）が、記述されるようにしてもよい。この記述は、例えば、ViewpointのDescriptorType elementを活用して新たにオブジェクト I D 情報（audio0bj）を定義するためのScheme（urn:mpeg:DASH:audio0bj:2013）を拡張することにより、行われる。この場合、動画再生端末 1 4 は、画像用の「AdaptationSet」に記述されたオブジェクト I D に対応するオブジェクトの音声ファイルを選択し、取得して再生する。

[0112] また、オブジェクト単位で音声ファイルが生成されるのではなく、全てのオブジェクトの符号化データが 1 本のオーディオストリームに多重化され、1 つの音声ファイルが生成されるようにしてもよい。

[0113] この場合、図 1 6 に示すように、MPDファイルの音声用の「AdaptationSet」には、音声ファイル用の「Representation」が 1 つ設けられ、「Segment」に、全てのオブジェクトの符号化データを含む音声ファイル（audio0bje.mp4）のURL等が記述される。このとき、Viewpointを拡張することにより、音声ファイルに対応する全てのオブジェクトのオブジェクト I D (1, 2, 3, 4, 5) も記述される。

- [0114] また、この場合、図 17 に示すように、MPDファイルの「Media Segment」を参照して取得される音声ファイル（以下、適宜、音声メディアファイルともいう）のmdat boxには、サブサンプルとして、各オブジェクトの符号化データ（Audio object）が配置される。
- [0115] 具体的には、音声メディアファイルには、セグメントより短い任意の時間であるサブセグメント単位でデータが配置される。このサブセグメント単位のデータの位置は、sidx boxにより指定される。また、サブセグメント単位のデータは、moof boxとmdat boxにより構成される。mdat boxは、複数のサンプルにより構成されるが、そのサンプルの各サブサンプルとして、各オブジェクトの符号化データが配置される。
- [0116] また、音声メディアファイルのsidx boxの後には、サンプルの情報を記述するgsix boxが配置される。このように、サンプルの情報を記述するgsix boxが、moof boxとは別に設けられるので、動画再生端末 14 は、サンプルの情報を素早く取得することができる。
- [0117] gsix boxには、図 18 に示すように、このgsix boxで管理する 1 以上のサンプルまたはサブサンプルからなるSample group entryの種別を表すgrouping_typeが記述される。例えば、Sample group entryが、オブジェクト単位の符号化データのサブサンプルである場合、図 17 に示すように、Sample group entryの種別は「obja」である。音声メディアファイルには、複数のgrouping_typeのgsix boxが配置される。
- [0118] また、図 18 に示すように、gsix boxには、各Sample group entryのインデックス（entry_index）と、音声メディアファイル内の位置を表すデータ位置情報としてのバイトレンジ（range_size）とが記述される。なお、インデックス（entry_index）が 0 である場合、対応するバイトレンジは、moof boxのバイトレンジ（図 17 の例ではa1）を表している。
- [0119] MPDファイルの「Initialization Segment」を参照して取得される音声ファイル（以下、適宜、音声初期化ファイルともいう）には、各Sample group entryが、どのオブジェクトの符号化データのサブサンプルであるかを表す情報

が記述される。

[0120] 具体的には、図19に示すように、音声初期化ファイルのsbt1 boxのsample group description box(sgpd)のAudioObjectSampleGroupEntryと対応付けられる、mvex boxのtype assignment box (typa) を用いて、その情報が表される。

[0121] 即ち、図20のAに示すように、AudioObjectSampleGroupEntryには、サンプルに含まれる符号化データに対応するオブジェクトID (audio_object_id) が1つずつ記述される。例えば、図20のBに示すように、4つのAudioObjectSampleGroupEntryのそれぞれに、オブジェクトIDとして、1,2,3,4が記述される。

[0122] 一方、図21に示すように、type assignment boxには、AudioObjectSampleGroupEntryごとに、そのAudioObjectSampleGroupEntryに対応するSample group entryのパラメータ (grouping_type_parameter) として、インデックスが記述される。

[0123] 以上のように音声メディアファイルと音声初期化ファイルが構成される。従って、動画再生端末14が、表示領域内のオブジェクトとして選択されたオブジェクトの符号化データを取得する際、音声初期化ファイルのstbl boxから、選択されたオブジェクトのオブジェクトIDが記述されたAudioObjectSampleGroupEntryが検索される。そして、mvex boxから、検索されたAudioObjectSampleGroupEntryに対応するSample group entryのインデックスが読み出される。そして、音声ファイルのsidxからサブセグメント単位のデータの位置が読み出され、gsixから、読み出されたインデックスのSample group entryのバイトレンジが読み出される。そして、サブセグメント単位のデータの位置とバイトレンジとに基づいて、mdat内に配置されている符号化データが取得される。これにより、選択されたオブジェクトの符号化データが取得される。

[0124] なお、上述した説明では、Sample group entryのインデックスとAudioObjectSampleGroupEntryのオブジェクトIDが、mvex boxを介して対応付けられ

たが、直接対応付けられるようにしてもよい。この場合、Sample group entryのインデックスが、AudioObjectSampleGroupEntryに記述される。

[0125] また、音声ファイルが複数のtrackから構成される場合、sgpdをmvexに格納すると、track間でsgpdを共用できる。

[0126] (情報処理システムの第2の例の概要)

図22は、本開示を適用した情報処理システムの第2の例の概要を説明する図である。

[0127] なお、図22において、図3と同一のものには、同一の符号を付してある。

[0128] 図22の例では、図3の場合と同様に、動画コンテンツの画像は、5（縦）×7（横）のタイルに分割されており、動画コンテンツの音声として、オブジェクト#1乃至#8の音声取得されている。

[0129] この場合に、ユーザにより、2（縦）×3（横）個のタイルからなる表示領域31が指定されると、第2の例では、図22に示すように、その表示領域31が、動画コンテンツの画像のサイズと同一のサイズに変換(拡大)され、表示画像111とされる。そして、表示画像111におけるオブジェクト#1乃至#8の位置に基づいて、オブジェクト#1乃至#8の音声合成され、表示画像111とともに出力される。即ち、表示領域31内のオブジェクト#1、オブジェクト#2、およびオブジェクト#6だけでなく、表示領域31外のオブジェクト#3乃至#5、オブジェクト#7、およびオブジェクト#8の音声も出力される。

[0130] (ストリーミング再生部の構成例)

本開示を適用した情報処理システムの第2の例の構成は、ストリーミング再生部の構成を除いて、図1の情報処理システム10の構成と同一であるので、以下では、ストリーミング再生部についてのみ説明する。

[0131] 図23は、本開示を適用した情報処理システムのストリーミング再生部の構成例を示すブロック図である。

[0132] 図23に示す構成のうち、図13の構成と同じ構成には同じ符号を付して

ある。重複する説明については適宜省略する。

[0133] 図23のストリーミング再生部120の構成は、MPD処理部92、音声合成処理部97、画像合成処理部101の代わりに、MPD処理部121、音声合成処理部123、画像合成処理部124が設けられる点、および、位置決定部122が新たに設けられる点が図13のストリーミング再生部90の構成と異なる。

[0134] ストリーミング再生部120のMPD処理部121は、MPD取得部91から供給されるMPDファイルから音声メタファイル用の「Segment」に記述されるURL等の情報を抽出し、メタファイル取得部93に供給する。また、MPD処理部121は、MPDファイルから画像用の「AdaptationSet」に記述される動画コンテンツの画像の画枠サイズ情報（以下、コンテンツ画枠サイズ情報という）を抽出し、位置決定部122に供給する。MPD処理部121は、MPDファイルから全てのオブジェクトの音声ファイル用の「Segment」に記述されるURL等の情報を抽出し、音声ファイル取得部95に供給する。

[0135] MPD処理部121は、MPDファイルから画像用の「AdaptationSet」に記述されるタイル位置情報を抽出し、画像選択部98に供給する。MPD処理部121は、MPDファイルから、画像選択部98から要求されるタイルの画像ファイル用の「Segment」に記述されるURL等の情報を抽出し、画像選択部98に供給する。

[0136] 位置決定部122は、メタファイル取得部93により取得された音声メタファイルに含まれるオブジェクト位置情報と、MPD処理部121から供給されるコンテンツ画枠サイズ情報を取得する。また、位置決定部122は、ユーザにより指定される表示領域の画枠サイズ情報である表示領域画枠サイズ情報を取得する。位置決定部122は、オブジェクト位置情報、コンテンツ画枠サイズ情報、および表示領域画枠サイズ情報に基づいて、表示画像におけるオブジェクトの位置を決定（認識）する。位置決定部122は、決定されたオブジェクトの位置を音声合成処理部123に供給する。

[0137] 音声合成処理部123は、位置決定部122から供給されるオブジェクト

の位置に基づいて、音声復号処理部 96 から供給されるオブジェクト単位の音声データを合成する。具体的には、音声合成処理部 123 は、オブジェクトの位置と音声を出力する各スピーカの位置とに基づいて、オブジェクトごとに、各スピーカに割り当てる音声データを決定する。そして、音声合成処理部 123 は、スピーカごとに、各オブジェクトの音声データを合成し、各スピーカの音声データとして出力する。オブジェクトの位置に基づいて各オブジェクトの音声データを合成する方法の詳細は、例えば、Ville Pulkki, “Virtual Sound Source Positioning Using Vector Base Amplitude Panning”, Journal of AES, vol.45, no.6, pp.456–466, 1997に記載されている。

[0138] 画像合成処理部 124 は、画像復号処理部 100 から供給されるタイル単位の画像データを合成する。画像合成処理部 124 は、変換部として機能し、合成後の画像データに対応する画像のサイズを動画コンテンツのサイズに変換して、表示画像を生成する。画像合成処理部 124 は、表示画像を出力する。

[0139] (オブジェクトの位置の決定方法の説明)

図 24 乃至図 26 は、図 23 の位置決定部 122 によるオブジェクトの位置の決定方法を説明する図である。

[0140] 表示画像 111 は、動画コンテンツ内の表示領域 31 を抽出し、その表示領域 31 のサイズを動画コンテンツのサイズに変換することにより生成される。従って、表示画像 111 は、図 24 に示すように、表示領域 31 の中心 C を表示画像 111 の中心 C' に移動させた後、図 25 に示すように、表示領域 31 のサイズを動画コンテンツのサイズに変換したものと等価である。

[0141] よって、まず、位置決定部 122 は、以下の式 (1) により、表示領域 31 の中心 O を表示画像 111 の中心 O' に移動させるときの水平方向の移動量 θ_{shift} を算出する。

[0142] [数1]

$$\theta_{\text{shift}} = \frac{\theta_{v1'} + \theta_{v2'} - \theta_{v1} - \theta_{v2}}{2} \quad \dots (1)$$

[0143] 式(1)において、 θ_{v1}' は、表示領域31の表示領域画枠サイズ情報のうちの左端の水平角度であり、 θ_{v2}' は、右端の水平角度である。また、 θ_{v1} は、コンテンツ画枠サイズ情報のうちの左端の水平角度であり、 θ_{v2} は、右端の水平角度である。

[0144] 次に、位置決定部122は、移動量 θ_{shift} を用いて、以下の式(2)により、表示領域31の中心0を表示画像111の中心0'に移動させた後の表示領域31の左端の水平角度 θ_{v1_shift}' と右端の水平角度 θ_{v2_shift}' を求める。

[0145] [数2]

$$\begin{aligned}\theta_{v1_shift}' &= \text{mod}(\theta_{v1}' + \theta_{shift} + 180^\circ, 360^\circ) - 180^\circ \\ \theta_{v2_shift}' &= \text{mod}(\theta_{v2}' + \theta_{shift} + 180^\circ, 360^\circ) - 180^\circ \quad \dots (2)\end{aligned}$$

[0146] 式(2)によれば、水平角度 θ_{v1_shift}' と水平角度 θ_{v2_shift}' は、 -180° から 180° までの範囲を超えないように求められる。

[0147] なお、上述したように、表示画像111は、表示領域31の中心0を表示画像111の中心0'に移動させた後、表示領域31のサイズを動画コンテンツのサイズに変換したものと等価である。従って、水平角度 θ_{v1} および θ_{v2} については、以下の式(3)が成立する。

[0148] [数3]

$$\begin{aligned}\theta_{v1} &= \frac{\theta_{v1} - \theta_{v2}}{\theta_{v1_shift}' - \theta_{v2_shift}'} * \theta_{v1_shift}' \\ \theta_{v2} &= \frac{\theta_{v1} - \theta_{v2}}{\theta_{v1_shift}' - \theta_{v2_shift}'} * \theta_{v2_shift}' \quad \dots (3)\end{aligned}$$

[0149] 位置決定部122は、以上のようにして移動量 θ_{shift} 、水平角度 θ_{v1_shift}' 、および水平角度 θ_{v2_shift}' を求めた後、表示画像111におけるオブジェクトの水平角度を求める。具体的には、位置決定部122は、以下の式(4)により、移動量 θ_{shift} を用いて、表示領域31の中心0を表示画像111の中心0'に移動させた後のオブジェクト#iの水平角度 θ_{Ai_shift} を求める。

[0150] [数4]

$$\theta_{Ai_shift} = \text{mod}(\theta_{Ai} + \theta_{shift} + 180^\circ, 360^\circ) - 180^\circ \quad \dots (4)$$

[0151] 式(4)において、 θ_{Ai} は、オブジェクト# i のオブジェクト位置情報のうちの水平角度である。また、式(4)によれば、水平角度 θ_{Ai_shift} は、 -180° から 180° までの範囲を超えないように求められる。

[0152] 次に、オブジェクト# i が表示領域31内に存在する場合、即ち $\theta_{v2_shift}' < \theta_{Ai_shift} < \theta_{v1_shift}'$ である場合、位置決定部122は、以下の式(5)により、表示画像111におけるオブジェクト# i の水平角度 θ_{Ai}' を求める。

[0153] [数5]

$$\theta_{Ai}' = \frac{\theta_{v1} - \theta_{v2}}{\theta_{v1_shift}' - \theta_{v2_shift}'} \cdot \left(\theta_{Ai_shift} - \frac{\theta_{v1} + \theta_{v2}}{2} \right) \quad \dots (5)$$

[0154] 式(5)によれば、表示領域31のサイズと表示画像111のサイズの比に応じて、表示画像111におけるオブジェクト# i の位置と表示画像111の中心C'との距離を拡大することにより、水平角度 θ_{Ai}' が求められる。

[0155] 一方、オブジェクト# i が表示領域31内に存在しない場合、即ち、 $-180^\circ \leq \theta_{Ai_shift} \leq \theta_{v2_shift}'$ 、または、 $\theta_{v1_shift}' \leq \theta_{Ai_shift} \leq 180^\circ$ である場合、位置決定部122は、以下の式(6)により、表示画像111におけるオブジェクト# i の水平角度 θ_{Ai}' を求める。

[0156] [数6]

$$\theta_{Ai}' = \frac{\theta_{v2} + 180^\circ}{\theta_{v2_shift}' + 180^\circ} \cdot (\theta_{Ai_shift} + 180^\circ) - 180^\circ$$

($-180^\circ \leq \theta_{Ai_shift} \leq \theta_{v2_shift}'$ の場合)

$$\theta_{Ai}' = \frac{180^\circ - \theta_{v1}}{180^\circ - \theta_{v1_shift}'} \cdot (\theta_{Ai_shift} - 180^\circ) + 180^\circ$$

($\theta_{v1_shift}' \leq \theta_{Ai_shift} \leq 180^\circ$ の場合)

... (6)

[0157] 式(6)によれば、図26に示すように、オブジェクト# i が表示領域31の右側の位置151に存在する($-180^\circ \leq \theta_{Ai_shift} \leq \theta_{v2_shift}'$)場合、角度R1と角度R2の比に応じて水平角度 θ_{Ai_shift} を拡大することにより、水平角度 θ_{Ai}' が求められる。なお、角度R1は、表示画像111の右端から視聴

者 1 5 3 の真後ろの位置 1 5 4 までの角度であり、角度 R 2 は、中心の移動後の表示領域 3 1 の右端から位置 1 5 4 までの角度である。

[0158] また、式 (6) によれば、オブジェクト # i が表示領域 3 1 の左側の位置 1 5 5 に存在する ($\theta_{v1_shift} \leq \theta_{Ai_shift} \leq 180^\circ$) 場合、角度 R 3 と角度 R 4 の比に応じて水平角度 θ_{Ai_shift} を拡大することにより、水平角度 θ_{Ai} が求められる。なお、角度 R 3 は、表示画像 1 1 1 の左端から位置 1 5 4 までの角度であり、角度 R 4 は、中心の移動後の表示領域 3 1 の左端から位置 1 5 4 までの角度である。

[0159] また、位置決定部 1 2 2 は、水平角度 θ_{Ai} と同様に、垂直角度 γ_{Ai} を求める。即ち、以下の式 (7) により、表示領域 3 1 の中心 C を表示画像 1 1 1 の中心 C' に移動させるときの垂直方向の移動量 γ_{shift} を算出する。

[0160] [数7]

$$\gamma_{shift} = \frac{\gamma_{v1'} + \gamma_{v2'} - \gamma_{v1} - \gamma_{v2}}{2} \quad \dots (7)$$

[0161] 式 (7) において、 $\gamma_{v1'}$ は、表示領域 3 1 の表示領域画枠サイズ情報のうちの上端の垂直角度であり、 $\gamma_{v2'}$ は、下端の垂直角度である。また、 γ_{v1} は、コンテンツ画枠サイズ情報のうちの上端の垂直角度であり、 γ_{v2} は、下端の垂直角度である。

[0162] 次に、位置決定部 1 2 2 は、移動量 γ_{shift} を用いて、以下の式 (8) により、表示領域 3 1 の中心 C を表示画像 1 1 1 の中心 C' に移動させた後の表示領域 3 1 の上端の垂直角度 γ_{v1_shift} と下端の垂直角度 γ_{v2_shift} を求める。

[0163] [数8]

$$\begin{aligned} \gamma_{v1_shift}' &= \text{mod}(\gamma_{v1'} + \gamma_{shift} + 90^\circ, 180^\circ) - 90^\circ \\ \gamma_{v2_shift}' &= \text{mod}(\gamma_{v2'} + \gamma_{shift} + 90^\circ, 180^\circ) - 90^\circ \quad \dots (8) \end{aligned}$$

[0164] 式 (8) によれば、垂直角度 γ_{v1_shift}' と垂直角度 γ_{v2_shift}' は、 -90° から 90° までの範囲を超えないように求められる。

[0165] 位置決定部 1 2 2 は、以上のようにして移動量 γ_{shift} 、垂直角度 γ_{v1_shift}' 、および垂直角度 γ_{v2_shift}' を求めた後、表示画像 1 1 1 におけるオブジェクト

の位置を求める。具体的には、位置決定部 1 2 2 は、以下の式 (9) により、移動量 γ_{shift} を用いて、表示領域 3 1 の中心 C を表示画像 1 1 1 の中心 C' に移動させた後のオブジェクト # i の垂直角度 $\gamma_{\text{Ai_shift}}$ を求める。

[0166] [数9]

$$\gamma_{\text{Ai_shift}} = \text{mod}(\gamma_{\text{Ai}} + \gamma_{\text{shift}} + 90^\circ, 180^\circ) - 90^\circ \quad \dots (9)$$

[0167] 式 (9) において、 γ_{Ai} は、オブジェクト # i のオブジェクト位置情報のうちの垂直角度である。また、式 (9) によれば、垂直角度 $\gamma_{\text{Ai_shift}}$ は、 -90° から 90° までの範囲を超えないように求められる。

[0168] 次に、位置決定部 1 2 2 は、以下の式 (10) により、表示画像 1 1 1 におけるオブジェクト # i の垂直角度 γ_{Ai}' を求める。

[0169] [数10]

$$\gamma_{\text{Ai}}' = \begin{cases} \frac{\gamma_{\text{v2}} + 90^\circ}{\gamma_{\text{v2_shift}}' + 90^\circ} \cdot (\gamma_{\text{Ai_shift}} + 90^\circ) - 90^\circ & (-90^\circ \leq \gamma_{\text{Ai_shift}} \leq \gamma_{\text{v2_shift}}' \text{ の場合}) \\ \frac{\gamma_{\text{v1}} - \gamma_{\text{v2}}}{\gamma_{\text{v1_shift}}' - \gamma_{\text{v2_shift}}'} \cdot \left(\gamma_{\text{Ai_shift}} - \frac{\gamma_{\text{v1}} + \gamma_{\text{v2}}}{2} \right) & (\gamma_{\text{v2_shift}}' < \gamma_{\text{Ai_shift}} < \gamma_{\text{v1_shift}}' \text{ の場合}) \\ \frac{90^\circ - \gamma_{\text{v1}}}{90^\circ - \gamma_{\text{v1_shift}}'} \cdot (\gamma_{\text{Ai_shift}} - 90^\circ) + 90^\circ & (\gamma_{\text{v1_shift}}' \leq \gamma_{\text{Ai_shift}} \leq 90^\circ \text{ の場合}) \end{cases} \quad \dots (10)$$

[0170] また、位置決定部 1 2 2 は、表示画像 1 1 1 におけるオブジェクト # i の距離 r_{Ai}' を、オブジェクト # i のオブジェクト位置情報のうちの距離 r_{Ai} に決定する。位置決定部 1 2 2 は、以上のようにして求められるオブジェクト # i の水平角度 θ_{Ai}' 、垂直角度 γ_{Ai}' 、および距離 r_{Ai} を、オブジェクト # i の位置として、音声合成処理部 1 2 3 に供給する。

[0171] 図 2 7 は、水平角度 θ_{Ai} と水平角度 θ_{Ai}' の関係を表す図である。

[0172] 図27のグラフにおいて、横軸は、水平角度 θ_{Ai} を表し、縦軸は、水平角度 θ_{Ai}' を表す。

[0173] 図27に示すように、 $\theta_{v2}' < \theta_{Ai} < \theta_{v1}'$ である場合、水平角度 θ_{Ai} は、移動量 θ_{shift} だけ移動され、拡大されることにより、水平角度 θ_{Ai}' になる。また、 $-180^\circ \leq \theta_{Ai} \leq \theta_{v2}'$ 、または、 $\theta_{v1}' \leq \theta_{Ai} \leq 180^\circ$ である場合、水平角度 θ_{Ai} は、移動量 θ_{shift} だけ移動され、縮小されることにより、水平角度 θ_{Ai}' になる。

[0174] (ストリーミング再生部の処理の説明)

図28は、図23のストリーミング再生部120のストリーミング再生処理を説明するフローチャートである。

[0175] 図28のステップS131において、ストリーミング再生部120のMPD取得部91は、Webサーバ12からMPDファイルを取得し、MPD処理部121に供給する。

[0176] ステップS132において、MPD処理部121は、MPD取得部91から供給されるMPDファイルから、画像用の「AdaptationSet」に記述されるコンテンツ画枠サイズ情報とタイル位置情報を取得する。MPD処理部121は、画枠サイズ情報を位置決定部122に供給し、タイル位置情報を画像選択部98に供給する。また、MPD処理部121は、音声メタファイル用の「Segment」に記述されるURL等の情報を抽出し、メタファイル取得部93に供給する。

[0177] ステップS133において、メタファイル取得部93は、MPD処理部121から供給されるURL等の情報に基づいて、そのURLで特定される音声メタファイルをWebサーバ12に要求し、取得する。メタファイル取得部93は、音声メタファイルに含まれるオブジェクト位置情報を位置決定部122に供給する。

[0178] ステップS134において、位置決定部122は、オブジェクト位置情報、コンテンツ画枠サイズ情報、および表示領域画枠サイズ情報に基づいて、表示画像におけるオブジェクトの位置を決定する位置決定処理を行う。この位置決定処理の詳細は、後述する図29を参照して説明する。

- [0179] ステップS 1 3 5において、MPD処理部 1 2 1は、MPDファイルから全てのオブジェクトの音声ファイル用の「Segment」に記述されるURL等の情報を抽出し、音声ファイル取得部 9 5に供給する。
- [0180] ステップS 1 3 6において、音声ファイル取得部 9 5は、MPD処理部 1 2 1から供給されるURL等の情報に基づいて、そのURLで特定される全てのオブジェクトの音声ファイルをWebサーバ 1 2に要求し、取得する。音声ファイル取得部 9 5は、取得されたオブジェクト単位の音声ファイルを音声復号処理部 9 6に供給する。
- [0181] ステップS 1 3 7乃至S 1 4 0の処理は、図 1 4のステップS 3 6乃至S 3 9の処理と同様であるので、説明は省略する。
- [0182] ステップS 1 4 1において、音声合成処理部 1 2 3は、位置決定部 1 2 2から供給されるオブジェクトの位置に基づいて、音声復号処理部 9 6から供給されるオブジェクト単位の音声データを合成し、出力する。
- [0183] ステップS 1 4 2において、画像合成処理部 1 2 4は、画像復号処理部 1 0 0から供給されるタイル単位の画像データを合成する。
- [0184] ステップS 1 4 3において、画像合成処理部 1 2 4は、合成後の画像データに対応する画像のサイズを動画コンテンツのサイズに変換し、表示画像を生成する。そして、画像合成処理部 1 2 4は、表示画像を出力し、処理を終了する。
- [0185] 図 2 9は、図 2 8のステップS 1 3 4の位置決定処理の詳細を説明するフローチャートである。この位置決定処理は、例えば、オブジェクトごとに行われる。
- [0186] 図 2 9のステップS 1 5 1において、位置決定部 1 2 2は、表示画像における水平角度 θ_{Ai} を推定する水平角度 θ_{Ai} 推定処理を行う。この水平角度 θ_{Ai} 推定処理の詳細は、後述する図 3 0を参照して説明する。
- [0187] ステップS 1 5 2において、位置決定部 1 2 2は、表示画像における垂直角度 γ_{Ai} を推定する垂直角度 γ_{Ai} 推定処理を行う。この垂直角度 γ_{Ai} 推定処理の詳細は、水平方向が垂直方向に代わる点を除いて、ステップS 1 5 1

の水平角度 θ_{Ai} を推定処理と同様であるので、詳細な説明は省略する。

[0188] ステップS 1 5 3において、位置決定部 1 2 2は、表示画像における距離 r_{Ai} を、メタファイル取得部 9 3から供給されるオブジェクト位置情報のうちの距離 r_{Ai} に決定する。

[0189] ステップS 1 5 4において、位置決定部 1 2 2は、水平角度 θ_{Ai} 、垂直角度 γ_{Ai} 、および距離 r_{Ai} を、オブジェクト # i の位置として音声合成処理部 1 2 3に出力する。そして、処理は、図 2 8のステップS 1 3 4に戻り、ステップS 1 3 5に進む。

[0190] 図 3 0は、図 2 9のステップS 1 5 1の水平角度 θ_{Ai} 推定処理の詳細を説明するフローチャートである。

[0191] 図 3 0のステップS 1 7 1において、位置決定部 1 2 2は、メタファイル取得部 9 3から供給されるオブジェクト位置情報のうちの水平角度 θ_{Ai} を取得する。

[0192] ステップS 1 7 2において、位置決定部 1 2 2は、MPD処理部 1 2 1から供給されるコンテンツ画枠サイズ情報と、ユーザにより指定される表示領域画枠サイズ情報を取得する。

[0193] ステップS 1 7 3において、位置決定部 1 2 2は、コンテンツ画枠サイズ情報と表示領域画枠サイズ情報に基づいて、上述した式 (1) により、移動量 θ_{shift} を算出する。

[0194] ステップS 1 7 4において、位置決定部 1 2 2は、移動量 θ_{shift} と表示領域画枠サイズを用いて、上述した式 (2) により、水平角度 $\theta_{v1 shift}$ および $\theta_{v2 shift}$ を算出する。

[0195] ステップS 1 7 5において、位置決定部 1 2 2は、水平角度 θ_{Ai} と移動量 θ_{shift} を用いて、上述した式 (4) により、水平角度 $\theta_{Ai shift}$ を求める。

[0196] ステップS 1 7 6において、位置決定部 1 2 2は、オブジェクト # i が表示領域 3 1内に存在する（オブジェクト # i の水平角度が表示領域 3 1の両端の水平角度の間である）かどうか、即ち $\theta_{v2 shift} < \theta_{Ai shift} < \theta_{v1 shift}$ であるかどうかを判定する。

- [0197] ステップS 1 7 6でオブジェクト# i が表示領域 3 1 内に存在すると判定された場合、即ち $\theta_{v2\ shift}^{\wedge} < \theta_{Ai\ shift} < \theta_{v1\ shift}^{\wedge}$ である場合、処理はステップS 1 7 7に進む。ステップS 1 7 7において、位置決定部 1 2 2は、コンテンツ画枠サイズ情報、水平角度 $\theta_{v1\ shift}^{\wedge}$ および $\theta_{v2\ shift}^{\wedge}$ 、並びに水平角度 $\theta_{Ai\ shift}$ に基づいて、上述した式(5)により水平角度 $\theta_{Ai}^{\wedge}$ を算出する。
- [0198] 一方、ステップS 1 7 6でオブジェクト# i が表示領域 3 1 内に存在しないと判定された場合、即ち、 $-180^{\circ} \leq \theta_{Ai\ shift} \leq \theta_{v2\ shift}^{\wedge}$ 、または、 $\theta_{v1\ shift}^{\wedge} \leq \theta_{Ai\ shift} \leq 180^{\circ}$ である場合、処理はステップS 1 7 8に進む。ステップS 1 7 8において、位置決定部 1 2 2は、コンテンツ画枠サイズ情報、水平角度 $\theta_{v1\ shift}^{\wedge}$ または $\theta_{v2\ shift}^{\wedge}$ 、および水平角度 $\theta_{Ai\ shift}$ に基づいて、上述した式(6)により水平角度 $\theta_{Ai}^{\wedge}$ を算出する。
- [0199] ステップS 1 7 7またはステップS 1 7 8の処理後、処理は、図 2 9のステップS 1 5 1に戻り、処理はステップS 1 5 2に進む。
- [0200] なお、第2の例では、表示画像のサイズが動画コンテンツのサイズと同一であるようにしたが、異なってもよい。
- [0201] また、第2の例において、全てのオブジェクトの音声データが合成されて出力されるのではなく、一部のオブジェクト(例えば、表示領域内のオブジェクト、表示領域から所定の範囲内のオブジェクト等)の音声データのみが合成されて出力されるようにしてもよい。出力する音声データのオブジェクトの選択方法は、予め決定されていてもよいし、ユーザにより指定されるようにしてもよい。
- [0202] さらに、上述した説明では、音声データは、オブジェクト単位の音声データのみであったが、音声データには、チャンネルオーディオの音声データ、H OA (Higher-Order Ambisonics) オーディオの音声データ、SAOC(Spatial Audio Object Coding)の音声データ、音声データのメタデータ(scene情報、動的・静的metadata)などが含まれるようにしてもよい。この場合、例えば、サブサンプルとして、各オブジェクトの符号化データだけでなく、これらのデータの符号化データも配置される。

[0203] <第1実施の形態>

(3Dオーディオファイルフォーマットの概要)

本開示を適用した第1実施の形態を説明する前に、まず、図31を参照して、MP4の3Dオーディオファイルフォーマットのトラックの概要を説明する。

[0204] MP4ファイルでは、トラックごとに、動画コンテンツのコーデック情報やファイル内の位置を示す位置情報を管理することができる。MP4の3Dオーディオファイルフォーマットでは、3Dオーディオ(Channel audio/Object audio/HOA audio/metadata)のオーディオストリーム(ES(Elementary Stream))の全てが、サンプル(フレーム)単位で1つのトラックとして記録される。また、3Dオーディオのコーデック情報(Profile/level/audio configuration)が、サンプルエントリ(sample entry)として格納される。

[0205] 3Dオーディオを構成するChannel audioは、チャンネル単位の音声データであり、Object audioは、オブジェクト単位の音声データであり、HOA audioは、球状の音声データであり、metadataは、Channel audio/Object audio/HOA audioのメタデータである。ここでは、Object audioは、オブジェクト単位の音声データであるものとするが、SAOCの音声データであってもよい。

[0206] (moovボックスの構造)

図32は、MP4ファイルのmoovボックスの構造を示す図である。

[0207] 図32に示すように、MP4ファイルでは、画像データと音声データが異なるトラックとして記録される。図32では、音声データのトラックの詳細は記述していないが、画像データのトラックと同様である。sample entryは、moovボックス内のstsdボックスに配置されるsample descriptionに含まれる。

[0208] ところで、放送やローカルストレージ再生では、Webサーバ側が、すべてのオーディオストリームを送出し、動画再生端末(クライアント)側は、すべてのオーディオストリーム(stream)をパースしながら、必要な3Dオーディオのオーディオストリームを復号(decode)し、出力(rendering)する。

ビットレート (Bitrate) が高い場合や、ローカルストレージの読み込みのrateに制約がある場合、必要な3Dオーディオのオーディオストリームのみを取得することにより、decode処理の負荷を軽減することが望まれている。

[0209] また、ストリーム再生では、動画再生端末（クライアント）側で、必要な3Dオーディオの符号化データだけを取得することにより、再生環境に最適な符号化速度のオーディオストリームを取得可能にすることが望まれている。

[0210] 従って、本開示では、3Dオーディオの符号化データを種類に応じてトラックに分割して音声ファイルに配置することにより、所定の種類の符号化データのみを効率良く取得可能にする。これにより、放送やローカルストレージ再生では、システムの負荷を軽減することができる。また、ストリーム再生では、帯域に応じて、必要な3Dオーディオの符号化データの最も品質の高いものを再生することができる。さらに、音声ファイル内の3Dオーディオのオーディオストリームの位置情報をサブセグメントのトラック単位で記録すれば済むので、オブジェクト単位の符号化データがサブサンプルに配置される場合に比べて、位置情報の情報量を削減することができる。

[0211] （トラックの概要）

図33は、本開示を適用した第1実施の形態におけるトラックの概要を説明する図である。

[0212] 図33に示すように、第1実施の形態では、3Dオーディオを構成するChannel audio/Object audio/HOA audio/metadataが、それぞれ、異なるトラック(Channel audio track/Object audio track(s)/HOA audio track/Object metadata track)のオーディオストリームとされる。object metadata trackには、オーディオメタデータのオーディオストリームが配置される。

[0213] また、3Dオーディオ全体に関する情報を配置するためのトラックとして、ベーストラック (Base Track) が設けられる。図33のベーストラックには、3Dオーディオ全体に関する情報がサンプルエントリに配置されるが、サンプルとしては何も配置されない。また、Base track, Channel audio trac

k, Object audio track(s), HOA audio track、およびObject metadata trackは、同一の音声ファイル (3daudio.mp4) として記録される。

[0214] Track Referenceは、例えばtrackボックスに配置され、対応するトラックの他のトラックとの参照関係を表す。具体的には、Track Referenceは、参照関係にある他のトラックのトラックに固有のID (以下、トラックIDという) を表す。図33の例では、Base track, Channel audio track, HOA audio track, Object metadata track、Object audio track(s)のトラックIDが、1, 2, 3, 4, 10...となっている。また、Base trackのTrack Referenceは2, 3, 4, 10...であり、Channel audio track/HOA audio track/Object metadata track/Object audio track(s)のTrack Referenceは、Base trackのトラックIDである1である。

[0215] 従って、Base trackと、Channel audio track/HOA audio track/Object metadata track/Object audio track(s)とは、参照関係にある。即ち、Base trackは、Channel audio track/HOA audio track/Object metadata track/Object audio track(s)の再生時に参照される。

[0216] (ベーストラックのサンプルエントリのシンタクスの例)

図34は、図33のベーストラックのサンプルエントリのシンタクスの例を示す図である。

[0217] 図34のconfigurationVersion, MPEGHAudioProfile、およびMPEGHAudioLevelは、3Dオーディオ全体に関する情報として、それぞれ、3Dオーディオのオーディオストリーム全体 (通常の3Dオーディオのオーディオストリーム) のconfig情報、profile情報、level情報を表す。また、図34のwidthとheightは、3Dオーディオ全体に関する情報として、それぞれ、動画コンテンツの水平方向の画素数、垂直方向の画素数を表す。theta1, theta2, gamma1、およびgamma2は、3Dオーディオ全体に関する情報として、それぞれ、動画コンテンツの画枠サイズ情報のうちの画枠の左端の水平角度 θ_{v1} 、右端の水平角度 θ_{v2} 、画枠の上端の垂直角度 γ_{v1} 、下端の垂直角度 γ_{v2} を表す。

[0218] (チャンネルオーディオトラックのサンプルエントリのシンタクスの例)

図35は、図33のチャンネルオーディオトラック (Channel audio track) のサンプルエントリのシンタクスの例を示す図である。

[0219] 図35のconfigurationVersion, MPEGHAudioProfile、およびMPEGHAudioLevelは、それぞれ、Channel Audioのconfig情報、profile情報、level情報を表す。

[0220] (オブジェクトオーディオトラックのサンプルエントリのシンタクスの例)

図36は、図33のオブジェクトオーディオトラック (Object audio track) のサンプルエントリのシンタクスの例を示す図である。

[0221] 図36のconfigurationVersion, MPEGHAudioProfile、およびMPEGHAudioLevelは、それぞれ、オブジェクトオーディオトラックに含まれる1以上のObject audioのconfig情報、profile情報、level情報を表す。object_is_fixedは、オブジェクトオーディオトラックに含まれる1以上のObject audioのオブジェクトが固定されているかどうかを表す。object_is_fixedが1である場合、オブジェクトが固定されていることを表し、0である場合、オブジェクトが移動することを表す。mpeg3daConfigは、オブジェクトオーディオトラックに含まれる1以上のObject audioのオブジェクトの識別情報のconfigを表す。

[0222] また、objectTheta1/objectTheta2/objectGamma1/objectGamma2/objectRengthは、オブジェクトオーディオトラックに含まれる1以上のObject audioのオブジェクト情報を表す。このオブジェクト情報は、Object_is_fixed=1の時に有効な情報である。

[0223] maxobjectTheta1, maxobjectTheta2, maxobjectGamma1, maxobjectGamma2、およびmaxobjectRengthは、オブジェクトオーディオトラックに含まれる1以上のObject audioのオブジェクトが移動する場合のオブジェクト情報の最大値を表す。

[0224] (H0Aオーディオトラックのサンプルエントリのシンタクスの例)

図37は、図33のH0Aオーディオトラック (H0A audio track) のサンプ

ルエントリのシンタクスの例を示す図である。

[0225] 図37のconfigurationVersion、MPEGHAudioProfile、およびMPEGHAudioLevelは、H0A audioのconfig情報、profile情報、level情報を表す。

[0226] (オブジェクトメタデータトラックのサンプルエントリのシンタクスの例)

図38は、図33のオブジェクトメタデータトラック (Object metadata track) のサンプルエントリのシンタクスの例を示す図である。

[0227] 図38のconfigurationVersionは、metadataのconfig情報を表す。

[0228] (3Dオーディオの音声ファイルのセグメント構造の第1の例)

図39は、本開示を適用した第1実施の形態における3Dオーディオの音声ファイルのセグメント構造の第1の例を示す図である。

[0229] 図39のセグメント構造では、Initial segmentが、ftypボックスとmoovボックスにより構成される。moovボックスには、音声ファイルに含まれるトラックごとにtrak boxが配置される。また、moovボックスには、各トラックのトラックIDと、media segment内のssixボックスで用いられるlevelとの対応関係を表す情報などを含むmvexボックスが配置される。

[0230] また、media segmentは、sidxボックス、ssixボックス、および、1以上のsubsegmentにより構成される。sidxボックスには、各subsegmentの音声ファイル内の位置を示す位置情報が配置される。ssixボックスには、mdatボックスに配置される各レベルのオーディオストリームの位置情報が含まれる。なお、レベルは、トラックに対応するものである。また、最初のトラックの位置情報は、moofボックスと最初のトラックのオーディオストリームからなるデータの位置情報である。

[0231] subsegmentは、任意の時間長ごとに設けられ、subsegmentには、全てのトラックに共通の1組のmoofボックスとmdatボックスのペアが設けられる。mdatボックスには、全てのtrackのオーディオストリームが、任意の時間長分だけまとめて配置され、moofボックスには、そのオーディオストリームの管理情報が配置される。mdatボックスに配置される各trackのオーディオストリー

ムは、track毎に連続している。

[0232] 図39の例では、トラックIDが1であるTrack1はbase Trackであり、トラックIDが2乃至NであるTrack2乃至trackNは、それぞれ、Channel Audio Track, Object audio track(s), HOA audio track, object metadata trackである。このことは、後述する図40においても同様である。

[0233] (3Dオーディオの音声ファイルのセグメント構造の第2の例)

図40は、本開示を適用した第1実施の形態における3Dオーディオの音声ファイルのセグメント構造の第2の例を示す図である。

[0234] 図40のセグメント構造は、トラックごとにmoofボックスとmdatボックスが設けられる点が、図39のセグメント構造と異なっている。

[0235] 即ち、図40のInitial segmentは、図39のInitial segmentと同様である。また、図40のmedia segmentは、図39のmedia segmentと同様に、sidxボックス、ssixボックス、および、1以上のsubsegmentにより構成される。sidxボックスには、図39のsidxボックスと同様に、各subsegmentの位置情報が配置される。ssixボックスには、moofボックスとmdatボックスからなる各レベルのデータの位置情報が含まれる。

[0236] subsegmentは、任意の時間長ごとに設けられ、subsegmentには、トラックごとにmoofボックスとmdatボックスのペアが設けられる。即ち、各トラックのmdatボックスには、そのトラックのオーディオストリームが、任意の時間長分だけまとめて配置（インターリーブ格納）され、moofボックスには、そのオーディオストリームの管理情報が配置される。

[0237] 図39および図40に示すように、各トラックのオーディオストリームは、任意の時間長分まとめて配置されるため、サンプル単位でまとめて配置される場合に比べて、HTTP等を介したオーディオストリームの取得効率が改善する。

[0238] (mvexボックスの記述例)

図41は、図39および図40のmvexボックスに配置されるlevel assignmentボックスの記述例を示す図である。

[0239] level assignmentボックスは、各トラックのトラックIDとssixボックスで用いられるレベル（level）を対応付けるボックスである。図41の例では、トラックIDが1であるベーストラックがレベル0に対応付けられ、トラックIDが2であるチャンネルオーディオトラックがレベル1に対応付けられている。また、トラックIDが3であるH0Aオーディオトラックがレベル2に対応付けられ、トラックIDが4であるオブジェクトメタデータトラックがレベル3に対応付けられている。さらに、トラックIDが10であるオブジェクトオーディオトラックがレベル4に対応付けられている。

[0240] （MPDファイルの記述例）

図42は、本開示を適用した第1実施の形態におけるMPDファイルの記述例を示す図である。

[0241] 図42に示すように、MPDファイルには、3Dオーディオの音声ファイル(3daudio.mp4)のセグメントを管理する「Representation」、そのセグメントに含まれるトラックを管理する「SubRepresentation」等が記述される。

[0242] 「Representation」と「SubRepresentation」には、対応するセグメントまたはトラックのcodecの種類を、3D audio file formatで定義されるコードで表す「codecs」が含まれる。また、「Representation」には、「id」、「associationId」、および「associationType」が含まれる。

[0243] 「id」は、それを含む「Representation」のIDである。「associationId」は、対応するトラックと他のトラックとの参照関係を表す情報であり、参照トラックの「id」である。「associationType」は、参照トラックとの参照関係（依存関係）の意味を表すコードであり、例えばMP4のtrack referenceの値と同じものが用いられる。

[0244] また、「SubRepresentation」には、対応するトラックと対応するレベルを表す値としてlevel assignmentボックスで設定されている値である「level」が含まれる。「SubRepresentation」には、参照関係を有する（依存する）他のトラック（以下、参照トラックという）に対応するレベルを表す値である「dependencyLevel」が含まれる。

- [0245] さらに、「SubRepresentation」には、3Dオーディオの選択に必要な情報として、`<EssentialProperty schemeIdUri= "urn:mpeg:DASH:3daudio:2014" value= " audioType, contentkind ,priority" >`が含まれる。
- [0246] また、Object audio trackの「SubRepresentation」には、`<EssentialProperty schemeIdUri= "urn:mpeg:DASH:viewingAngle:2014" value= " $\theta, \gamma, r$ " >`が含まれる。この θ, γ, r は、「SubRepresentation」に対応するオブジェクトが固定されている場合、それぞれ、オブジェクト位置情報のうちの水平角度、垂直角度、距離である。一方、オブジェクトが移動する場合、この θ, γ, r は、それぞれ、オブジェクト位置情報の最大値のうちの水平角度の最大値、垂直角度の最大値、距離の最大値である。
- [0247] 図43は、図42のEssential Propertyの定義を示す図である。
- [0248] 図43の左上側は、`<EssentialProperty schemeIdUri= "urn:mpeg:DASH:3daudio:2014" value= " audioType, contentkind ,priority" >`のAudioTypeの定義を示している。AudioTypeは、対応するトラックの3Dオーディオの種類を表す。
- [0249] 図43の例では、AudioTypeが1である場合、対応するトラックの音声データが、3DオーディオのうちのChannel audioであることを表し、AudioTypeが2である場合、対応するトラックの音声データが、HOA audioであることを表す。また、AudioTypeが3である場合、対応するトラックの音声データが、Object audioであることを表し、AudioTypeが4である場合、対応するトラックの音声データが、metadataであることを表す。
- [0250] また、図43の右側は、`<EssentialProperty schemeIdUri= "urn:mpeg:DASH:3daudio:2014" value= " audioType, contentkind ,priority" >`のcontentkindの定義を示している。contentkindは、対応する音声の内容を表す。図43の例では、例えば、contentkindが3である場合、対応する音声は音楽(music)である。
- [0251] 図43の左下に示すように、Priorityは23008-3で定義されているものであり、対応するObjectの処理優先度を表す。Priorityとしては、オーディオ

ストリームの途中で変更されない場合にのみobjectの処理優先度を表す値が記述され、変更される場合には0が記述される。

[0252] (情報処理システムの概要)

図44は、本開示を適用した第1実施の形態における情報処理システムの概要を説明する図である。

[0253] 図44に示す構成のうち、図1の構成と同じ構成には同じ符号を付してある。重複する説明については適宜省略する。

[0254] 図44の情報処理システム140は、ファイル生成装置141と接続されるWebサーバ142と動画再生端末144が、インターネット13を介して接続されることにより構成される。

[0255] 情報処理システム140では、MPEG-DASHに準ずる方式で、Webサーバ142が、動画コンテンツのビデオストリームをタイル単位で動画再生端末144に配信（タイルドストリーミング）する。また、情報処理システム140では、Webサーバ142が、再生対象のタイルに対応するObject audio, Channel audio、またはHOA audioのオーディオストリームを動画再生端末144に配信する。

[0256] 情報処理システム140のファイル生成装置141は、音声ファイル生成部56が第1実施の形態における音声ファイルを生成し、MPD生成部57が第1実施の形態におけるMPDファイルを生成する点等を除いて、図11のファイル生成装置11と同様である。

[0257] 具体的には、ファイル生成装置141は、動画コンテンツの画像データを取得し、タイル単位で符号化してビデオストリームを生成する。ファイル生成装置141は、各タイルのビデオストリームを、セグメントごとにファイル化する。ファイル生成装置141は、その結果得られる各タイルの画像ファイルをWebサーバ142にアップロードする。

[0258] また、ファイル生成装置141は、動画コンテンツの3Dオーディオを取得し、3Dオーディオの種類（Channel audio/Object audio/HOA audio/metadata）ごとに符号化してオーディオストリームを生成する。ファイル生成装

置 1 4 1 は、オーディオストリームに対して、3 Dオーディオの種類ごとにトラックを割り当てる。ファイル生成装置 1 4 1 は、各トラックのオーディオストリームがサブセグメント単位で配置される図 3 9 または図 4 0 のセグメント構造の音声ファイルを生成し、Webサーバ 1 4 2 にアップロードする。

[0259] ファイル生成装置 1 4 1 は、画枠サイズ情報、タイル位置情報、オブジェクト位置情報等を含むMPDファイルを生成する。ファイル生成装置 1 4 1 は、MPDファイルをWebサーバ 1 4 2 にアップロードする。

[0260] Webサーバ 1 4 2 は、ファイル生成装置 1 4 1 からアップロードされた画像ファイル、音声ファイル、およびMPDファイルを格納する。

[0261] 図 4 4 の例では、Webサーバ 1 4 2 には、タイル # 1 の複数のセグメントの画像ファイルからなるセグメント群と、タイル # 2 の複数のセグメントの画像ファイルからなるセグメント群とが格納されている。また、Webサーバ 1 4 2 には、3 Dオーディオの音声ファイルからなるセグメント群が格納されている。

[0262] Webサーバ 1 4 2 は、動画再生端末 1 4 4 からの要求に応じて、格納している画像ファイル、音声ファイル、MPDファイル等を動画再生端末 1 4 4 に送信する。

[0263] 動画再生端末 1 4 4 は、制御用ソフトウェア 1 6 1、動画再生ソフトウェア 1 6 2、アクセス用ソフトウェア 1 6 3 などを実行する。

[0264] 制御用ソフトウェア 1 6 1 は、Webサーバ 1 4 2 からストリーミングするデータを制御するソフトウェアである。具体的には、制御用ソフトウェア 1 6 1 は、動画再生端末 1 4 4 にWebサーバ 1 4 2 からMPDファイルを取得させる。

[0265] また、制御用ソフトウェア 1 6 1 は、動画再生ソフトウェア 1 6 2 から指示される表示領域と、MPDファイルに含まれるタイル位置情報とに基づいて、表示領域内のタイルを特定する。そして、制御用ソフトウェア 1 6 1 は、そのタイルの画像ファイルの送信要求をアクセス用ソフトウェア 1 6 3 に指令する。

- [0266] 制御用ソフトウェア161は、Object audioを再生対象とする場合、音声ファイル内の画枠サイズ情報の送信要求をアクセス用ソフトウェア163に指令する。また、制御用ソフトウェア161は、metadataのオーディオストリームの送信要求をアクセス用ソフトウェア163に指令する。制御用ソフトウェア161は、その指令に応じてWebサーバ142から送信されてくる画枠サイズ情報およびmetadataのオーディオストリームに含まれるオブジェクト位置情報、並びに、表示領域に基づいて、表示領域内の画像に対応するオブジェクトを特定する。そして、制御用ソフトウェア161は、そのオブジェクトのオーディオストリームの送信要求をアクセス用ソフトウェア163に指令する。
- [0267] また、制御用ソフトウェア161は、Channel audioまたはHOA audioを再生対象とする場合、Channel audioまたはHOA audioのオーディオストリームの送信要求をアクセス用ソフトウェア163に指令する。
- [0268] 動画再生ソフトウェア162は、Webサーバ142から取得された画像ファイルと音声ファイルを再生するソフトウェアである。具体的には、動画再生ソフトウェア162は、ユーザにより表示領域が指定されると、その表示領域を制御用ソフトウェア161に指示する。また、動画再生ソフトウェア162は、その指示に応じてWebサーバ142から取得された画像ファイルと音声ファイルを復号する。動画再生ソフトウェア162は、復号の結果得られるタイル単位の画像データを合成して出力する。また、動画再生ソフトウェア162は、復号の結果得られるObject audio、Channel audio、またはHOA audioを必要に応じて合成し、出力する。
- [0269] アクセス用ソフトウェア163は、HTTPを用いたインターネット13を介したWebサーバ142との通信を制御するソフトウェアである。具体的には、アクセス用ソフトウェア163は、制御用ソフトウェア161の指令に応じて、画像ファイル、音声ファイル内の画枠サイズ情報や所定のオーディオストリームの送信要求を、動画再生端末144に送信させる。また、アクセス用ソフトウェア163は、その送信要求に応じて、Webサーバ142から送信

されてくる画像ファイル、音声ファイル内の画枠サイズ情報や所定のオーディオストリームを動画再生端末 144 に受信させる。

[0270] (ファイル生成装置の構成例)

図 45 は、図 44 のファイル生成装置 141 の構成例を示すブロック図である。

[0271] 図 45 に示す構成のうち、図 11 の構成と同じ構成には同じ符号を付してある。重複する説明については適宜省略する。

[0272] 図 45 のファイル生成装置 141 の構成は、音声符号化処理部 55、音声ファイル生成部 56、MPD生成部 57、サーバアップロード処理部 58 の代わりに、音声符号化処理部 171、音声ファイル生成部 172、MPD生成部 173、サーバアップロード処理部 174 が設けられる点が、図 11 のファイル生成装置 11 の構成と異なる。

[0273] 具体的には、ファイル生成装置 141 の音声符号化処理部 171 は、外部から入力される動画コンテンツの 3D オーディオの種類 (Channel audio/Object audio/HOA audio/metadata) ごとに符号化し、オーディオストリームを生成する。音声符号化処理部 171 は、3D オーディオの種類ごとのオーディオストリームを音声ファイル生成部 172 に供給する。

[0274] 音声ファイル生成部 172 は、音声符号化処理部 171 から供給されるオーディオストリームに対して、3D オーディオの種類ごとにトラックを割り当てる。音声ファイル生成部 172 は、各トラックのオーディオストリームがサブセグメント単位で配置される図 39 または図 40 のセグメント構造の音声ファイルを生成する。このとき、音声ファイル生成部 172 は、外部から入力される画枠サイズ情報をサンプルエントリに格納する。音声ファイル生成部 172 は、生成された音声ファイルを MPD生成部 173 に供給する。

[0275] MPD生成部 173 は、画像ファイル生成部 53 から供給される各タイルの画像ファイルを格納する Webサーバ 142 の URL 等を決定する。また、MPD生成部 173 は、音声ファイル生成部 172 から供給される音声ファイルを格納する Webサーバ 142 の URL 等を決定する。

- [0276] MPD生成部173は、画像情報生成部54から供給される画像情報をMPDファイルの画像用の「AdaptationSet」に配置する。また、MPD生成部173は、各タイルの画像ファイルのURL等を、そのタイルの画像ファイル用の「Representation」の「Segment」に配置する。
- [0277] MPD生成部173は、音声ファイルのURL等を、その音声ファイル用の「Representation」の「Segment」に配置する。また、MPD生成部173は、外部から入力される各オブジェクトのオブジェクト位置情報等を、そのオブジェクトのObject metadata track用の「Sub Representation」に配置する。MPD生成部173は、以上のようにして各種の情報が配置されたMPDファイル、並びに、画像ファイルおよび音声ファイルをサーバアップロード処理部174に供給する。
- [0278] サーバアップロード処理部174は、MPD生成部173から供給される各タイルの画像ファイル、音声ファイル、およびMPDファイルを、Webサーバ142にアップロードする。
- [0279] (ファイル生成装置の処理の説明)
- 図46は、図45のファイル生成装置141のファイル生成処理を説明するフローチャートである。
- [0280] 図46のステップS191乃至S195の処理は、図12のステップS11乃至S15の処理と同様であるので、説明は省略する。
- [0281] ステップS196において、音声符号化処理部171は、外部から入力される動画コンテンツの3Dオーディオを種類(Channel audio/Object audio/HOA audio/metadata)ごとに符号化し、オーディオストリームを生成する。音声符号化処理部171は、3Dオーディオの種類ごとのオーディオストリームを音声ファイル生成部172に供給する。
- [0282] ステップS197において、音声ファイル生成部172は、音声符号化処理部171から供給されるオーディオストリームに対して、3Dオーディオの種類ごとにトラックを割り当てる。
- [0283] ステップS198において、音声ファイル生成部172は、各トラックの

オーディオストリームがサブセグメント単位で配置される図39または図40のセグメント構造の音声ファイルを生成する。このとき、音声ファイル生成部172は、外部から入力される画枠サイズ情報をサンプルエントリに格納する。音声ファイル生成部172は、生成された音声ファイルをMPD生成部173に供給する。

[0284] ステップS199において、MPD生成部173は、画像情報生成部54から供給される画像情報、各ファイルのURL、オブジェクト位置情報等を含むMPDファイルを生成する。MPD生成部173は、画像ファイル、音声ファイル、およびMPDファイルをサーバアップロード処理部174に供給する。

[0285] ステップS200において、サーバアップロード処理部174は、MPD生成部173から供給される画像ファイル、音声ファイル、およびMPDファイルを、Webサーバ142にアップロードする。そして、処理は終了する。

[0286] (動画再生端末の機能的構成例)

図47は、図44の動画再生端末144が制御用ソフトウェア161、動画再生ソフトウェア162、およびアクセス用ソフトウェア163を実行することにより実現されるストリーミング再生部の構成例を示すブロック図である。

[0287] 図47に示す構成のうち、図13の構成と同じ構成には同じ符号を付してある。重複する説明については適宜省略する。

[0288] 図47のストリーミング再生部190の構成は、MPD処理部92、音声選択部94、音声ファイル取得部95、音声復号処理部96、音声合成処理部97の代わりに、MPD処理部191、音声選択部193、音声ファイル取得部192、音声復号処理部194、音声合成処理部195が設けられる点、および、メタファイル取得部93が設けられない点が、図13のストリーミング再生部90の構成と異なる。

[0289] ストリーミング再生部190は、選択されたオブジェクトの再生対象の音声データを取得する方法等が異なる点等を除いて、図13のストリーミング再生部90と同様である。

- [0290] 具体的には、ストリーミング再生部 190 のMPD処理部 191 は、MPD取得部 91 から供給されるMPDファイルから音声ファイル用の「Segment」に記述される再生対象のセグメントの音声ファイルのURL等の情報を抽出し、音声ファイル取得部 192 に供給する。
- [0291] MPD処理部 191 は、MPDファイルから画像用の「AdaptationSet」に記述されるタイル位置情報を抽出し、画像選択部 98 に供給する。MPD処理部 191 は、MPDファイルから、画像選択部 98 から要求されるタイルの画像ファイル用の「Segment」に記述されるURL等の情報を抽出し、画像選択部 98 に供給する。
- [0292] 音声ファイル取得部 192 は、Object audioを再生対象とする場合、MPD処理部 191 から供給されるURL等の情報に基づいて、そのURLで特定される音声ファイル内のBase trackのInitial SegmentをWebサーバ 142 に要求し、取得する。
- [0293] また、音声ファイル取得部 192 は、音声ファイルのURL等の情報に基づいて、そのURLで特定される音声ファイル内のobject metadata trackのオーディオストリームをWebサーバ 142 に要求し、取得する。音声ファイル取得部 192 は、object metadata trackのオーディオストリームに含まれるオブジェクト位置情報、Base trackのInitial Segmentに含まれる画枠サイズ情報、および音声ファイルのURL等の情報を音声選択部 193 に供給する。
- [0294] また、Channel audioを再生対象とする場合、音声ファイル取得部 192 は、音声ファイルのURL等の情報に基づいて、そのURLで特定される音声ファイル内のChannel audio trackのオーディオストリームをWebサーバ 142 に要求し、取得する。音声ファイル取得部 192 は、取得されたChannel audio trackのオーディオストリームを音声復号処理部 194 に供給する。
- [0295] HOA audioを再生対象とする場合、音声ファイル取得部 192 は、Channel audioを再生対象とする場合と同様の処理を行う。その結果、HOA audio trackのオーディオストリームが音声復号処理部 194 に供給される。
- [0296] なお、Object audio, Channel audio、およびHOA audioのいずれを再生対象

とするかは、例えば、ユーザからの指令により決定される。

[0297] 音声選択部 193 は、音声ファイル取得部 192 から供給される画枠サイズ情報とオブジェクト位置情報とに基づいて、各オブジェクトの画像上の位置を算出する。音声選択部 193 は、各オブジェクトの画像上の位置に基づいて、ユーザにより指定される表示領域内のオブジェクトを選択する。音声選択部 193 は、音声ファイル取得部 192 から供給される音声ファイルの URL 等の情報に基づいて、その URL で特定される音声ファイル内の、選択されたオブジェクトの Object audio track のオーディオストリームを Web サーバ 142 に要求し、取得する。音声選択部 193 は、取得された Object audio track のオーディオストリームを音声復号処理部 194 に供給する。

[0298] 音声復号処理部 194 は、音声ファイル取得部 192 から供給される Channel audio track または HOA audio track のオーディオストリーム、もしくは、音声選択部 193 から Object audio track のオーディオストリームを復号する。音声復号処理部 194 は、復号の結果得られる Channel audio, HOA audio、または Object audio を音声合成処理部 195 に供給する。

[0299] 音声合成処理部 195 は、音声復号処理部 194 から供給される Object audio, Channel audio、または HOA audio を必要に応じて合成し、出力する。

[0300] (動画再生端末の処理の説明)

図 48 は、図 47 のストリーミング再生部 190 のチャンネルオーディオ再生処理を説明するフローチャートである。このチャンネルオーディオ再生処理は、例えば、ユーザにより Channel audio が再生対象として選択される場合に行われる。

[0301] 図 48 のステップ S221 において、MPD 処理部 191 は、MPD 取得部 91 から供給される MPD ファイルを解析し、「SubRepresentation」に記述される essential property と codec に基づいて、再生対象のセグメントの Channel audio の「SubRepresentation」を特定する。また、MPD 処理部 191 は、MPD ファイルから、再生対象のセグメントの音声ファイル用の「Segment」に記述される URL 等の情報を抽出し、音声ファイル取得部 192 に供給する。

- [0302] ステップS 2 2 2において、MPD処理部 1 9 1は、ステップS 2 2 1で特定された「SubRepresentation」のdependencyLevelに基づいて、参照トラックであるBase trackのレベルを特定し、音声ファイル取得部 1 9 2に供給する。
- [0303] ステップS 2 2 3において、音声ファイル取得部 1 9 2は、MPD処理部 1 9 1から供給されるURL等の情報に基づいて、再生対象のセグメントのInitial SegmentをWebサーバ 1 4 2に要求し、取得する。
- [0304] ステップS 2 2 4において、音声ファイル取得部 1 9 2は、Initial Segment内のLevel assignmentボックスから、channel audio trackと参照トラックであるBase trackのレベルに対応するトラックIDを取得する。
- [0305] ステップS 2 2 5において、音声ファイル取得部 1 9 2は、channel audio trackと参照トラックであるBase trackのトラックIDに基づいて、Initial Segmentの、そのトラックIDに対応するtrakボックス内のサンプルエントリを取得する。音声ファイル取得部 1 9 2は、取得されたサンプルエントリに含まれるコーデック情報を音声復号処理部 1 9 4に供給する。
- [0306] ステップS 2 2 6において、音声ファイル取得部 1 9 2は、MPD処理部 1 9 1から供給されるURL等の情報に基づいて、Webサーバ 1 4 2に要求し、再生対象のセグメントの音声ファイルの先頭からsidxボックスとssixボックスを取得する。
- [0307] ステップS 2 2 7において、音声ファイル取得部 1 9 2は、ステップS 2 2 3で取得されたsidxボックスとssixボックスから、再生対象のサブセグメントのchannel audio trackと参照トラックの位置情報を取得する。なお、ここでは、参照トラックであるBase trackにはオーディオストリームが含まれないため、参照トラックの位置情報はない。
- [0308] ステップS 2 2 8において、音声ファイル取得部 1 9 2は、channel audio trackの位置情報と再生対象のセグメントの音声ファイルのURL等の情報に基づいて、mdatボックスに配置されるchannel audio trackのオーディオストリームをWebサーバ 1 4 2に要求し、取得する。音声ファイル取得部 1 9 2は、

取得されたchannel audio trackのオーディオストリームを音声復号処理部 194 に供給する。

[0309] ステップS 229において、音声復号処理部 194 は、音声ファイル取得部 192 から供給されるコーデック情報に基づいて、channel audio trackのオーディオストリームに対して復号を行う。音声ファイル取得部 192 は、その結果得られるchannel audioを音声合成処理部 195 に供給する。

[0310] ステップS 230において、音声合成処理部 195 は、channel audioを出力し、処理を終了する。

[0311] なお、図示は省略するが、ストリーミング再生部 190 によるH0A audioを再生するH0Aオーディオ再生処理は、図 48 のチャンネルオーディオ再生処理と同様に行われる。

[0312] 図 49 は、図 47 のストリーミング再生部 190 のオブジェクト特定処理を説明するフローチャートである。このオブジェクト特定処理は、例えば、ユーザによりObject audioが再生対象として選択されており、再生領域が変化したときに行われる。

[0313] 図 49 のステップS 251において、音声選択部 193 は、ユーザ操作などにより、ユーザにより指定された表示領域を取得する。

[0314] ステップS 252において、MPD処理部 191 は、MPD取得部 91 から供給されるMPDファイルを解析し、「SubRepresentation」に記述されるessential property とcodecに基づいて、再生対象のセグメントのmetadataの「SubRepresentation」を特定する。また、MPD処理部 191 は、MPDファイルから音声ファイル用の「Segment」に記述される再生対象のセグメントの音声ファイルのURL等の情報を抽出し、音声ファイル取得部 192 に供給する。

[0315] ステップS 253において、MPD処理部 191 は、ステップS 252で特定された「SubRepresentation」のdependencyLevelに基づいて、参照トラックであるBase trackのレベルを特定し、音声ファイル取得部 192 に供給する。

[0316] ステップS 254において、音声ファイル取得部 192 は、MPD処理部 19

1 から供給されるURL等の情報に基づいて、再生対象のセグメントのInitial SegmentをWebサーバ142に要求し、取得する。

[0317] ステップS255において、音声ファイル取得部192は、Initial Segment内のLevel assignmentボックスから、object metadata trackと参照トラックであるBase trackのレベルに対応するトラックIDを取得する。

[0318] ステップS256において、音声ファイル取得部192は、object metadata trackと参照トラックであるBase trackのトラックIDに基づいて、Initial Segmentの、そのトラックIDに対応するtrakボックス内のサンプルエントリを取得する。音声ファイル取得部192は、参照トラックであるBase trackのサンプルエントリに含まれる画枠サイズ情報を音声選択部193に供給する。また、音声ファイル取得部192は、Initial Segmentを音声選択部193に供給する。

[0319] ステップS257において、音声ファイル取得部192は、MPD処理部191から供給されるURL等の情報に基づいて、Webサーバ142に要求し、再生対象のセグメントの音声ファイルの先頭からsidxボックスとssixボックスを取得する。

[0320] ステップS258において、音声ファイル取得部192は、ステップS257で取得されたsidxボックスとssixボックスから、再生対象のサブセグメントのobject metadata trackと参照トラックの位置情報を取得する。なお、ここでは、参照トラックであるBase trackにはオーディオストリームが含まれないため、参照トラックの位置情報はない。音声ファイル取得部192は、sidxボックスとssixボックスを音声選択部193に供給する。

[0321] ステップS259において、音声ファイル取得部192は、object metadata trackの位置情報と再生対象のセグメントの音声ファイルのURL等の情報に基づいて、mdatボックスに配置されるobject metadata trackのオーディオストリームをWebサーバ142に要求し、取得する。

[0322] ステップS260において、音声ファイル取得部192は、ステップS256で取得されたサンプルエントリに含まれるコーデック情報に基づいて、

ステップS 2 5 9で取得されたobject metadata trackのオーディオストリームを復号する。音声ファイル取得部1 9 2は、復号の結果得られるmetadataに含まれるオブジェクト位置情報を音声選択部1 9 3に供給する。また、音声ファイル取得部1 9 2は、MPD処理部1 9 1から供給される音声ファイルのURL等の情報を音声選択部1 9 3に供給する。

[0323] ステップS 2 6 1において、音声選択部1 9 3は、音声ファイル取得部1 9 2から供給される画枠サイズ情報およびオブジェクト位置情報、並びに、ユーザにより指定される表示領域に基づいて、表示領域内のオブジェクトを選択する。そして、処理は終了する。

[0324] 図5 0は、図4 9のオブジェクト特定処理後にストリーミング再生部1 9 0により行われる特定オブジェクトオーディオ再生処理を説明するフローチャートである。

[0325] 図5 0のステップS 2 8 1において、MPD処理部1 9 1は、MPD取得部9 1から供給されるMPDファイルを解析し、「SubRepresentation」に記述されるessential property とcodecに基づいて、選択されたオブジェクトのobject audioの「SubRepresentation」を特定する。

[0326] ステップS 2 8 2において、MPD処理部1 9 1は、ステップS 2 8 1で特定された「SubRepresentation」のdependencyLevelに基づいて、参照トラックであるBase trackのレベルを特定し、音声ファイル取得部1 9 2に供給する。

[0327] ステップS 2 8 3において、音声ファイル取得部1 9 2は、Initial Segment内のLevel assignmentボックスから、object audio trackと参照トラックであるBase trackのレベルに対応するトラックIDを取得し、音声選択部1 9 3に供給する。

[0328] ステップS 2 8 4において、音声選択部1 9 3は、object audio trackと参照トラックであるBase trackのトラックIDに基づいて、Initial Segmentの、そのトラックIDに対応するtrakボックス内のサンプルエントリを取得する。このInitial Segmentは、図4 9のステップS 2 5 6で音声ファイル取

得部 192 から供給されたものである。音声選択部 193 は、取得されたサンプルエントリに含まれるコーデック情報を音声復号処理部 194 に供給する。

- [0329] ステップ S 285 において、音声選択部 193 は、ステップ S 258 で音声ファイル取得部 192 から供給される sidx ボックスと ssix ボックスから、再生対象のサブセグメントの選択されたオブジェクトの object audio track と参照トラックの位置情報を取得する。なお、ここでは、参照トラックである Base track にはオーディオストリームが含まれないため、参照トラックの位置情報はない。
- [0330] ステップ S 286 において、音声選択部 193 は、object audio track の位置情報と再生対象のセグメントの音声ファイルの URL 等の情報に基づいて、mdat ボックスに配置される、選択されたオブジェクトの object audio track のオーディオストリームを Web サーバ 142 に要求し、取得する。音声選択部 193 は、取得された object audio track のオーディオストリームを音声復号処理部 194 に供給する。
- [0331] ステップ S 287 において、音声復号処理部 194 は、音声選択部 193 から供給されるコーデック情報に基づいて、object audio track のオーディオストリームを復号する。音声選択部 193 は、復号の結果得られる object audio を音声合成処理部 195 に供給する。
- [0332] ステップ S 288 において、音声合成処理部 195 は、音声復号処理部 194 から供給される object audio を合成して出力する。そして、処理は終了する。
- [0333] 以上のように、情報処理システム 140 では、ファイル生成装置 141 が、3D オーディオが 3D オーディオの種類に応じて複数のトラックに分割されて配置される音声ファイルを生成する。そして、動画再生端末 144 は、その音声ファイルのうちの所定の種類の 3D オーディオのトラックのオーディオストリームを取得する。従って、動画再生端末 144 は、所定の種類の 3D オーディオのオーディオストリームを効率良く取得することができる。

よって、ファイル生成装置 141 は、所定の種類の 3D オーディオのオーディオストリームの取得効率を向上させる音声ファイルを生成しているといえる。

[0334] <第2実施の形態>

(トラックの概要)

図51は、本開示を適用した第2実施の形態におけるトラックの概要を説明する図である。

[0335] 図51に示すように、第2実施の形態では、Base trackのサンプルとして、ベースサンプルが記録される点が、第1実施の形態と異なる。ベースサンプルは、Channel audio/Object audio/HOA audio/metadataのsampleへの参照情報により構成される。ベースサンプルに含まれる参照情報により参照されるChannel audio/Object audio/HOA audio/metadataのサンプルを、参照情報の配置順に配置することにより、トラックに分割される前の3Dオーディオのオーディオストリームを生成することができる。

[0336] (ベーストラックのサンプルエントリのシンタクスの例)

図52は、図51のベーストラックのサンプルエントリのシンタクスの例を示す図である。

[0337] 図52のシンタクスは、サンプルエントリが、図33のBase trackのサンプルエントリであることを表す「mha1」の代わりに、図51のBase trackのサンプルエントリであることを表す「mha2」が記述される点を除いて、図34のシンタクスと同一である。

[0338] (ベースサンプルの構造例)

図53は、ベースサンプルの構造例を示す図である。

[0339] 図53に示すように、ベースサンプルは、サンプル単位のChannel audio/Object audio/HOA audio/metadataのextractorをサブサンプルとして構成される。Channel audio/Object audio/HOA audio/metadataのextractorは、extractorの種別、並びに、対応するChannel audio track/Object audio track(s)/HOA audio track/Object metadata trackのサブサンプルのoffsetおよびsi

zeにより構成される。このoffsetは、ベースサンプルのサブサンプルのファイル内の位置と、Channel audio track/Object audio track(s)/HOA audio track/Object metadata trackのサンプルのファイル内の位置の差分である。即ち、offsetは、それを含むベースサンプルのサブサンプルに対応する他のトラックのサンプルのファイル内の位置を示す情報である。

[0340] 図54は、ベースサンプルのシンタクスの例を示す図である。

[0341] 図54に示すように、ベースサンプルでは、Object audio trackのサンプルにおいてobject audioが格納されるSCEエレメントが、extractorを格納するEXTエレメントに入れ換えられる。

[0342] 図55は、extractorのデータの例を示す図である。

[0343] 図55に示すように、extractorには、extractorの種別、並びに、対応するChannel audio track/Object audio track(s)/HOA audio track/Object metadata trackのサブサンプルのoffsetおよびsizeが記述される。

[0344] なお、AVC(Advanced Video Coding)/HEVC(High Efficiency Video Coding)で定義している、NAL(Network Abstraction Layer)の構造を活用して、audio elementaryやconfig情報を格納できるように拡張してもよい。

[0345] 第2実施の形態における情報処理システムおよび情報処理システムによる処理は、第1実施の形態と同様であるので、説明は省略する。

[0346] <第3実施の形態>

(トラックの概要)

図56は、本開示を適用した第3実施の形態におけるトラックの概要を説明する図である。

[0347] 図56に示すように、第3実施の形態では、Base trackのサンプルとして、ベースサンプルとmetadataのサンプルが記録され、Object metadata trackが設けられない点が、第1実施の形態と異なっている。

[0348] 第3実施の形態における情報処理システムおよび情報処理システムによる処理は、オブジェクト位置情報を取得するために、Object metadata trackの代わりにBase trackのオーディオストリームが取得される点を除いて、第1

実施の形態と同様であるので、説明は省略する。

[0349] <第4実施の形態>

(トラックの概要)

図57は、本開示を適用した第4実施の形態におけるトラックの概要を説明する図である。

[0350] 図57に示すように、第4実施の形態では、各トラックが異なるファイル(3da_base.mp4/3da_channel.mp4/3da_object_1.mp4/3da_hoa.mp4/3da_meta.mp4)として記録される点が、第1実施の形態と異なっている。この場合、所望のトラックのファイルを、HTTPを介して取得することにより、所望のトラックの音声データのみを取得することができる。従って、HTTPを介した所望のトラックの音声データの取得を効率的に行うことができる。

[0351] (MPDファイルの記述例)

図58は、本開示を適用した第4実施の形態におけるMPDファイルの記述例を示す図である。

[0352] 図58に示すように、MPDファイルには、3Dオーディオの各音声ファイル(3da_base.mp4/3da_channel.mp4/3da_object_1.mp4/3da_hoa.mp4/3da_meta.mp4)のセグメントを管理する「Representation等が記述される。

[0353] 「Representation」には、「codecs」、「id」、「associationId」、および「associationType」が含まれる。また、Channel audio track/Object audio track(s)/HOA audio track/Object metadata trackの「Representation」には、「<EssentialProperty schemeIdUri=“urn:mpeg:DASH:3daudio:2014” value=“audioType, contentkind, priority”>」も含まれる。さらに、Object audio track(s)の「Representation」には、<EssentialProperty schemeIdUri=“urn:mpeg:DASH:viewingAngle:2014” value=“ θ , γ , r ”>が含まれる。

[0354] (情報処理システムの概要)

図59は、本開示を適用した第4実施の形態における情報処理システムの概要を説明する図である。

- [0355] 図59に示す構成のうち、図1の構成と同じ構成には同じ符号を付してある。重複する説明については適宜省略する。
- [0356] 図59の情報処理システム210は、ファイル生成装置211と接続されるWebサーバ212と動画再生端末214が、インターネット13を介して接続されることにより構成される。
- [0357] 情報処理システム210では、MPEG-DASHに準ずる方式で、Webサーバ212が、動画コンテンツのビデオストリームをタイル単位で動画再生端末214に配信（タイルドストリーミング）する。また、情報処理システム210では、Webサーバ212が、再生対象のタイルに対応するObject audio, Channel audio、またはHOA audioの音声ファイルを動画再生端末214に配信する。
- [0358] 具体的には、ファイル生成装置211は、動画コンテンツの画像データを取得し、タイル単位で符号化してビデオストリームを生成する。ファイル生成装置211は、各タイルのビデオストリームを、セグメントごとにファイル化する。ファイル生成装置211は、その結果得られる各タイルの画像ファイルをWebサーバ212にアップロードする。
- [0359] また、ファイル生成装置211は、動画コンテンツの3Dオーディオを取得し、3Dオーディオの種類（Channel audio/Object audio/HOA audio/meta data）ごとに符号化してオーディオストリームを生成する。ファイル生成装置211は、オーディオストリームに対して、3Dオーディオの種類ごとにトラックを割り当てる。ファイル生成装置211は、トラックごとに、オーディオストリームを配置した音声ファイルを生成し、Webサーバ212にアップロードする。
- [0360] ファイル生成装置211は、画枠サイズ情報、タイル位置情報、オブジェクト位置情報等を含むMPDファイルを生成する。ファイル生成装置211は、MPDファイルをWebサーバ212にアップロードする。
- [0361] Webサーバ212は、ファイル生成装置211からアップロードされた画像ファイル、3Dオーディオの種類ごとの音声ファイル、およびMPDファイルを

格納する。

- [0362] 図59の例では、Webサーバ212には、タイル#1の複数のセグメントの画像ファイルからなるセグメント群と、タイル#2の複数のセグメントの画像ファイルからなるセグメント群とが格納されている。また、Webサーバ212には、Channel audioの音声ファイルからなるセグメント群と、オブジェクト#1の音声ファイルからなるセグメント群とが格納されている。
- [0363] Webサーバ212は、動画再生端末214からの要求に応じて、格納している画像ファイル、3Dオーディオの所定の種類の音声ファイル、MPDファイル等を動画再生端末214に送信する。
- [0364] 動画再生端末214は、制御用ソフトウェア221、動画再生ソフトウェア222、アクセス用ソフトウェア223などを実行する。
- [0365] 制御用ソフトウェア221は、Webサーバ212からストリーミングするデータを制御するソフトウェアである。具体的には、制御用ソフトウェア221は、動画再生端末214にWebサーバ212からMPDファイルを取得させる。
- [0366] また、制御用ソフトウェア221は、動画再生ソフトウェア222から指示される表示領域と、MPDファイルに含まれるタイル位置情報とに基づいて、表示領域内のタイルを特定する。そして、制御用ソフトウェア221は、そのタイルの画像ファイルの送信要求をアクセス用ソフトウェア223に指令する。
- [0367] 制御用ソフトウェア221は、Object audioを再生対象とする場合、Base trackの音声ファイルの送信要求をアクセス用ソフトウェア223に指令する。また、制御用ソフトウェア221は、Object metadata trackの音声ファイルの送信要求をアクセス用ソフトウェア223に指令する。制御用ソフトウェア221は、その指令に応じてWebサーバ142から送信されてくるBase trackの音声ファイル内の画枠サイズ情報とmetadataの音声ファイルに含まれるオブジェクト位置情報を取得する。制御用ソフトウェア221は、画枠サイズ情報、オブジェクト位置情報、および表示領域に基づいて、表示領域内

の画像に対応するオブジェクトを特定する。そして、制御用ソフトウェア 221 は、そのオブジェクトの音声ファイルの送信要求をアクセス用ソフトウェア 223 に指令する。

[0368] また、制御用ソフトウェア 221 は、Channel audio または HOA audio を再生対象とする場合、Channel audio または HOA audio の音声ファイルの送信要求をアクセス用ソフトウェア 223 に指令する。

[0369] 動画再生ソフトウェア 222 は、Webサーバ 212 から取得された画像ファイルと音声ファイルを再生するソフトウェアである。具体的には、動画再生ソフトウェア 222 は、ユーザにより表示領域が指定されると、その表示領域を制御用ソフトウェア 221 に指示する。また、動画再生ソフトウェア 222 は、その指示に応じて Webサーバ 212 から取得された画像ファイルと音声ファイルを復号する。動画再生ソフトウェア 222 は、復号の結果得られるタイル単位の画像データを合成して出力する。また、動画再生ソフトウェア 222 は、復号の結果得られる Object audio、Channel audio、または HOA audio を必要に応じて合成し、出力する。

[0370] アクセス用ソフトウェア 223 は、HTTP を用いたインターネット 13 を介した Webサーバ 212 との通信を制御するソフトウェアである。具体的には、アクセス用ソフトウェア 223 は、制御用ソフトウェア 221 の指令に応じて、画像ファイルや所定の音声ファイルの送信要求を、動画再生端末 214 に送信させる。また、アクセス用ソフトウェア 223 は、その送信要求に応じて、Webサーバ 212 から送信されてくる画像ファイルや所定の音声ファイルを動画再生端末 214 に受信させる。

[0371] (ファイル生成装置の構成例)

図 60 は、図 59 のファイル生成装置 211 の構成例を示すブロック図である。

[0372] 図 60 に示す構成のうち、図 45 の構成と同じ構成には同じ符号を付してある。重複する説明については適宜省略する。

[0373] 図 60 のファイル生成装置 211 の構成は、音声ファイル生成部 172、M

PD生成部 173、サーバアップロード処理部 174 の代わりに、音声ファイル生成部 241、MPD生成部 242、サーバアップロード処理部 243 が設けられる点が、図 45 のファイル生成装置 141 の構成と異なる。

[0374] 具体的には、ファイル生成装置 211 の音声ファイル生成部 241 は、音声符号化処理部 171 から供給されるオーディオストリームに対して、3Dオーディオの種類ごとにトラックを割り当てる。音声ファイル生成部 241 は、トラックごとに、オーディオストリームを配置した音声ファイルを生成する。このとき、音声ファイル生成部 241 は、外部から入力される画枠サイズ情報を、Base trackの音声ファイルのサンプルエントリに格納する。音声ファイル生成部 241 は、3Dオーディオの種類ごとの音声ファイルをMPD生成部 242 に供給する。

[0375] MPD生成部 242 は、画像ファイル生成部 53 から供給される各タイルの画像ファイルを格納するWebサーバ 212 のURL等を決定する。また、MPD生成部 242 は、3Dオーディオの種類ごとに、音声ファイル生成部 241 から供給される音声ファイルを格納するWebサーバ 212 のURL等を決定する。

[0376] MPD生成部 242 は、画像情報生成部 54 から供給される画像情報をMPDファイルの画像用の「AdaptationSet」に配置する。また、MPD生成部 242 は、各タイルの画像ファイルのURL等を、そのタイルの画像ファイル用の「Representation」の「Segment」に配置する。

[0377] MPD生成部 242 は、3Dオーディオの種類ごとに、音声ファイルのURL等を、その音声ファイル用の「Representation」の「Segment」に配置する。また、MPD生成部 242 は、外部から入力される各オブジェクトのオブジェクト位置情報等を、そのオブジェクトのObject metadata track用の「Representation」に配置する。MPD生成部 242 は、以上のようにして各種の情報が配置されたMPDファイル、並びに、画像ファイルおよび3Dオーディオの種類ごとの音声ファイルを、サーバアップロード処理部 243 に供給する。

[0378] サーバアップロード処理部 243 は、MPD生成部 242 から供給される各タイルの画像ファイル、3Dオーディオの種類ごとの音声ファイル、およびMPD

ファイルを、Webサーバ212にアップロードする。

[0379] (ファイル生成装置の処理の説明)

図61は、図60のファイル生成装置211のファイル生成処理を説明するフローチャートである。

[0380] 図61のステップS301乃至S307の処理は、図46のステップS191乃至S197の処理と同様であるので、説明は省略する。

[0381] ステップS308において、音声ファイル生成部241は、トラックごとに、オーディオストリームが配置された音声ファイルを生成する。このとき、音声ファイル生成部241は、外部から入力される画枠サイズ情報をBase trackの音声ファイル内のサンプルエントリに格納する。音声ファイル生成部241は、生成された3Dオーディオの種類ごとの音声ファイルをMPD生成部242に供給する。

[0382] ステップS309において、MPD生成部242は、画像情報生成部54から供給される画像情報、各ファイルのURL、オブジェクト位置情報等を含むMPDファイルを生成する。MPD生成部242は、画像ファイル、3Dオーディオの種類ごとの音声ファイル、およびMPDファイルをサーバアップロード処理部243に供給する。

[0383] ステップS310において、サーバアップロード処理部243は、MPD生成部242から供給される画像ファイル、3Dオーディオの種類ごとの音声ファイル、およびMPDファイルを、Webサーバ212にアップロードする。そして、処理は終了する。

[0384] (動画再生端末の機能的構成例)

図62は、図59の動画再生端末214が制御用ソフトウェア221、動画再生ソフトウェア222、およびアクセス用ソフトウェア223を実行することにより実現されるストリーミング再生部の構成例を示すブロック図である。

[0385] 図62に示す構成のうち、図13や図47の構成と同じ構成には同じ符号を付してある。重複する説明については適宜省略する。

[0386] 図62のストリーミング再生部260の構成は、MPD処理部92、メタファイル取得部93、音声選択部94、音声ファイル取得部95、音声復号処理部96、音声合成処理部97の代わりに、MPD処理部261、メタファイル取得部262、音声選択部263、音声ファイル取得部264、音声復号処理部194、音声合成処理部195が設けられる点が、図13のストリーミング再生部90の構成と異なる。

[0387] 具体的には、ストリーミング再生部260のMPD処理部261は、Object audioを再生対象とする場合、MPD取得部91から供給されるMPDファイルから、再生対象のセグメントのobject metadata trackの音声ファイルの「Segment」に記述されるURL等の情報を抽出し、メタファイル取得部262に供給する。また、MPD処理部261は、MPDファイルから、音声選択部263から要求されるオブジェクトのobject audio trackの音声ファイルの「Segment」に記述されるURL等の情報を抽出し、音声選択部263に供給する。さらに、MPD処理部261は、MPDファイルから、再生対象のセグメントのBase trackの音声ファイルの「Segment」に記述されるURL等の情報を抽出し、メタファイル取得部262に供給する。

[0388] また、MPD処理部261は、Channel audioまたはHOA audioを再生対象とする場合、MPDファイルから、再生対象のセグメントのChannel audio trackまたはHOA audio trackの音声ファイルの「Segment」に記述されるURL等の情報を抽出する。MPD処理部261は、そのURL等の情報を、音声選択部263を介して音声ファイル取得部264に供給する。

[0389] なお、Object audio, Channel audio、およびHOA audioのいずれを再生対象とするかは、例えば、ユーザからの指令により決定される。

[0390] MPD処理部261は、MPDファイルから画像用の「AdaptationSet」に記述されるタイル位置情報を抽出し、画像選択部98に供給する。MPD処理部261は、MPDファイルから、画像選択部98から要求されるタイルの画像ファイル用の「Segment」に記述されるURL等の情報を抽出し、画像選択部98に供給する。

[0391] メタファイル取得部262は、MPD処理部261から供給されるURL等の情報に基づいて、そのURLで特定されるobject metadata trackの音声ファイルをWebサーバ212に要求し、取得する。メタファイル取得部93は、object metadata trackの音声ファイルに含まれるオブジェクト位置情報を音声選択部263に供給する。

[0392] また、メタファイル取得部262は、音声ファイルのURL等の情報に基づいて、そのURLで特定されるBase trackの音声ファイルのInitial SegmentをWebサーバ142に要求し、取得する。メタファイル取得部262は、Initial Segmentのサンプルエントリに含まれる画枠サイズ情報を音声選択部263に供給する。

[0393] 音声選択部263は、メタファイル取得部262から供給される画枠サイズ情報とオブジェクト位置情報とに基づいて、各オブジェクトの画像上の位置を算出する。音声選択部263は、各オブジェクトの画像上の位置に基づいて、ユーザにより指定される表示領域内のオブジェクトを選択する。音声選択部263は、選択されたオブジェクトのobject audio trackの音声ファイルのURL等の情報をMPD処理部261に要求する。音声選択部263は、その要求に応じてMPD処理部261から供給されるURL等の情報を音声ファイル取得部264に供給する。

[0394] 音声ファイル取得部264は、音声選択部263から供給される、object audio track, Channel audio track、またはHOA audio trackの音声ファイルのURL等の情報に基づいて、そのURLで特定される音声ファイルのオーディオストリームをWebサーバ12に要求し、取得する。音声ファイル取得部95は、取得されたオブジェクト単位の音声ファイルを音声復号処理部194に供給する。

[0395] (動画再生端末の処理の説明)

図63は、図62のストリーミング再生部260のチャンネルオーディオ再生処理を説明するフローチャートである。このチャンネルオーディオ再生処理は、例えば、ユーザによりChannel audioが再生対象として選択される場

合に行われる。

- [0396] 図63のステップS331において、MPD処理部261は、MPD取得部91から供給されるMPDファイルを解析し、「Representation」に記述されるessential property とcodecに基づいて、再生対象のセグメントのChannel audioの「Representation」を特定する。また、MPD処理部261は、その「Representation」に含まれる「Segment」に記述される再生対象のセグメントのChannel audio trackの音声ファイルのURL等の情報を抽出し、音声選択部263を介して音声ファイル取得部264に供給する。
- [0397] ステップS332において、MPD処理部261は、ステップS331で特定された「Representation」のassociationIdに基づいて、参照トラックであるBase trackの「Representation」を特定する。MPD処理部261は、その「Representation」に含まれる「Segment」に記述される参照トラックの音声ファイルのURL等の情報を抽出し、音声選択部263を介して音声ファイル取得部264に供給する。
- [0398] ステップS333において、音声ファイル取得部264は、音声選択部263から供給されるURL等の情報に基づいて、再生対象のセグメントのChannel audio trackおよび参照トラックの音声ファイルのInitial SegmentをWebサーバ212に要求し、取得する。
- [0399] ステップS334において、音声ファイル取得部264は、取得されたInitial Segmentのtrakボックス内のサンプルエントリを取得する。音声ファイル取得部264は、取得されたサンプルエントリに含まれるコーデック情報を音声復号処理部194に供給する。
- [0400] ステップS335において、音声ファイル取得部264は、音声選択部263から供給されるURL等の情報に基づいて、Webサーバ142に要求し、再生対象のセグメントのChannel audio trackの音声ファイルの先頭からsidxボックスとssixボックスを取得する。
- [0401] ステップS336において、音声ファイル取得部264は、ステップS333で取得されたsidxボックスとssixボックスから、再生対象のサブセグメ

ントの位置情報を取得する。

- [0402] ステップS 3 3 7において、音声選択部2 6 3は、ステップS 3 3 7で取得された位置情報と再生対象のセグメントのchannel audio trackの音声ファイルのURL等の情報に基づいて、その音声ファイル内のmdatボックスに配置されるchannel audio trackのオーディオストリームをWebサーバ1 4 2に要求し、取得する。音声選択部2 6 3は、取得されたchannel audio trackのオーディオストリームを音声復号処理部1 9 4に供給する。
- [0403] ステップS 3 3 8において、音声復号処理部1 9 4は、音声ファイル取得部2 6 4から供給されるコーデック情報に基づいて、音声選択部2 6 3から供給されるchannel audio trackのオーディオストリームに対して復号を行う。音声選択部2 6 3は、その結果得られるchannel audioを音声合成処理部1 9 5に供給する。
- [0404] ステップS 3 3 9において、音声合成処理部1 9 5は、channel audioを出力し、処理を終了する。
- [0405] なお、図示は省略するが、ストリーミング再生部2 6 0によるH0A audioを再生するH0Aオーディオ再生処理は、図6 3のチャンネルオーディオ再生処理と同様に行われる。
- [0406] 図6 4は、図6 2のストリーミング再生部2 6 0のオブジェクトオーディオ再生処理のを説明するフローチャートである。このオブジェクトオーディオ再生処理は、例えば、ユーザによりObject audioが再生対象として選択されており、再生領域が変化したときに行われる。
- [0407] 図6 4のステップS 3 5 1において、音声選択部2 6 3は、ユーザ操作などにより、ユーザにより指定された表示領域を取得する。
- [0408] ステップS 3 5 2において、MPD処理部2 6 1は、MPD取得部9 1から供給されるMPDファイルを解析し、「Representation」に記述されるessential property とcodecに基づいて、再生対象のセグメントのmetadataの「Representation」を特定する。また、MPD処理部2 6 1は、その「Representation」に含まれる「Segment」に記述される再生対象のセグメントのobject metadata

trackの音声ファイルのURL等の情報を抽出し、メタファイル取得部262に供給する。

- [0409] ステップS353において、MPD処理部261は、ステップS352で特定された「Representation」のassociationIdに基づいて、参照トラックであるBase trackの「Representation」を特定する。MPD処理部261は、その「Representation」に含まれる「Segment」に記述される参照トラックの音声ファイルのURL等の情報を抽出し、メタファイル取得部262に供給する。
- [0410] ステップS354において、メタファイル取得部262は、MPD処理部261から供給されるURL等の情報に基づいて、再生対象のセグメントのobject metadata trackおよび参照トラックの音声ファイルのInitial SegmentをWebサーバ212に要求し、取得する。
- [0411] ステップS355において、メタファイル取得部262は、取得されたInitial Segmentのtrakボックス内のサンプルエントリを取得する。メタファイル取得部262は、参照トラックであるBase trackのサンプルエントリに含まれる画枠サイズ情報を音声ファイル取得部264に供給する。
- [0412] ステップS356において、メタファイル取得部262は、MPD処理部261から供給されるURL等の情報に基づいて、Webサーバ142に要求し、再生対象のセグメントのobject metadata trackの音声ファイルの先頭からsidxボックスとssixボックスを取得する。
- [0413] ステップS357において、メタファイル取得部262は、ステップS356で取得されたsidxボックスとssixボックスから、再生対象のサブセグメントの位置情報を取得する。
- [0414] ステップS358において、メタファイル取得部262は、ステップS357で取得された位置情報と再生対象のセグメントのobject metadata trackの音声ファイルのURL等の情報に基づいて、その音声ファイル内のmdatボックスに配置されるobject metadata trackのオーディオストリームをWebサーバ142に要求し、取得する。
- [0415] ステップS359において、メタファイル取得部262は、ステップS3

55で取得されたサンプルエントリに含まれるコーデック情報に基づいて、ステップS358で取得されたobject metadata trackのオーディオストリームを復号する。メタファイル取得部262は、復号の結果得られるmetadataに含まれるオブジェクト位置情報を音声選択部263に供給する。

[0416] ステップS360において、音声選択部263は、メタファイル取得部262から供給される画枠サイズ情報およびオブジェクト位置情報、並びに、ユーザにより指定される表示領域に基づいて、表示領域内のオブジェクトを選択する。音声選択部263は、選択されたオブジェクトのobject audio trackの音声ファイルのURL等の情報をMPD処理部261に要求する。

[0417] ステップS361において、MPD処理部261は、MPD取得部91から供給されるMPDファイルを解析し、「Representation」に記述されるessential property とcodecに基づいて、選択されたオブジェクトのobject audioの「Representation」を特定する。また、MPD処理部261は、その「Representation」に含まれる「Segment」に記述される、再生対象のセグメントの選択されたオブジェクトのobject audio trackの音声ファイルのURL等の情報を抽出し、音声選択部263を介して音声ファイル取得部264に供給する。

[0418] ステップS362において、MPD処理部261は、ステップS361で特定された「Representation」のassociationIdに基づいて、参照トラックであるBase trackの「Representation」を特定する。MPD処理部261は、その「Representation」に含まれる「Segment」に記述される参照トラックの音声ファイルのURL等の情報を抽出し、音声選択部263を介して音声ファイル取得部264に供給する。

[0419] ステップS363において、音声ファイル取得部264は、音声選択部263から供給されるURL等の情報に基づいて、再生対象のセグメントのobject audio trackおよび参照トラックの音声ファイルのInitial SegmentをWebサーバ212に要求し、取得する。

[0420] ステップS364において、音声ファイル取得部264は、取得されたInitial Segmentのtrakボックス内のサンプルエントリを取得する。音声ファイ

ル取得部264は、そのサンプルエントリに含まれるコーデック情報を音声復号処理部194に供給する。

[0421] ステップS365において、音声ファイル取得部264は、音声選択部263から供給されるURL等の情報に基づいて、Webサーバ142に要求し、再生対象のセグメントのobject audio trackの音声ファイルの先頭からsidxボックスとssixボックスを取得する。

[0422] ステップS366において、音声ファイル取得部264は、ステップS365で取得されたsidxボックスとssixボックスから、再生対象のサブセグメントの位置情報を取得する。

[0423] ステップS367において、音声ファイル取得部264は、ステップS366で取得された位置情報と再生対象のセグメントのobject audio trackの音声ファイルのURL等の情報に基づいて、その音声ファイル内のmdatボックスに配置されるobject audio trackのオーディオストリームをWebサーバ142に要求し、取得する。音声ファイル取得部264は、取得されたobject audio trackのオーディオストリームを音声復号処理部194に供給する。

[0424] ステップS368およびS369の処理は、図50のステップS287およびS288の処理と同様であるので、説明は省略する。

[0425] なお、上述した説明では、音声選択部263は、表示領域内の全てのオブジェクトを選択したが、表示領域内のオブジェクトのうちの処理優先度が高いもののみを選択したり、所定の内容の音声のオブジェクトのみを選択するようにしてもよい。

[0426] 図65は、音声選択部263が表示領域内のオブジェクトのうちの処理優先度が高いもののみを選択する場合のオブジェクトオーディオ再生処理を説明するフローチャートである。

[0427] 図65のオブジェクトオーディオ再生処理は、図64のステップS360の代わりに図65のステップS390の処理が行われる点を除いて、図64のオブジェクトオーディオ再生処理と同様である。即ち、図65のステップS381乃至S389およびS391乃至S399の処理は、図64のステ

ップS 3 5 1乃至S 3 5 9およびS 3 6 1乃至S 3 6 9の処理と同様である。従って、以下では、ステップS 3 9 0の処理についてのみ説明する。

[0428] 図6 5のステップS 3 9 0において、音声ファイル取得部2 6 4は、画枠サイズ情報、オブジェクト位置情報、表示領域、および、各オブジェクトのpriorityに基づいて、表示領域内の処理優先度が高いオブジェクトを選択する。具体的には、音声ファイル取得部2 6 4は、画枠サイズ情報、オブジェクト位置情報、および表示領域に基づいて、表示領域内のオブジェクトを特定する。そして、音声ファイル取得部2 6 4は、特性されたオブジェクトのうちの、priorityが所定値以上であるオブジェクトを選択する。なお、priorityは、例えば、MPD処理部2 6 1がMPDファイルを解析することにより、特定されたオブジェクトのobject audioの「Representation」から取得される。音声選択部2 6 3は、選択されたオブジェクトのobject audio trackの音声ファイルのURL等の情報をMPD処理部2 6 1に要求する。

[0429] 図6 6は、音声選択部2 6 3が表示領域内のオブジェクトのうちの、処理優先度が高い所定の内容の音声のオブジェクトのみを選択する場合のオブジェクトオーディオ再生処理を説明するフローチャートである。

[0430] 図6 6のオブジェクトオーディオ再生処理は、図6 4のステップS 3 6 0の代わりに図6 6のステップS 4 2 0の処理が行われる点を除いて、図6 4のオブジェクトオーディオ再生処理と同様である。即ち、図6 6のステップS 3 8 1乃至S 3 8 9およびS 3 9 1乃至S 3 9 9の処理は、図6 4のステップS 4 1 1乃至S 4 1 9およびS 4 2 1乃至S 4 2 9の処理と同様である。従って、以下では、ステップS 4 2 0の処理についてのみ説明する。

[0431] 図6 6のステップS 4 2 0において、音声ファイル取得部2 6 4は、画枠サイズ情報、オブジェクト位置情報、表示領域、並びに、各オブジェクトのpriorityおよびcontentkindに基づいて、表示領域内の処理優先度が高い所定の内容の音声のオブジェクトを選択する。具体的には、音声ファイル取得部2 6 4は、画枠サイズ情報、オブジェクト位置情報、および表示領域に基づいて、表示領域内のオブジェクトを特定する。そして、音声ファイル取得部

264は、特性されたオブジェクトのうちの、priorityが所定値以上であり、かつ、contentkindが所定値であるオブジェクトを選択する。

[0432] なお、priorityおよびcontentkindは、例えば、MPD処理部261がMPDファイルを解析することにより、特定されたオブジェクトのobject audioの「Representation」から取得される。音声選択部263は、選択されたオブジェクトのobject audio trackの音声ファイルのURL等の情報をMPD処理部261に要求する。

[0433] 図67は、priority（プライオリティ）に基づいて選択されたオブジェクトの例を示す図である。

[0434] 図67の例では、オブジェクト#1(object1)乃至オブジェクト#4(object4)が、表示領域内のオブジェクトであり、表示領域内のオブジェクトのうちのpriorityが2以下であるオブジェクトが選択される。なお、priorityは、数字が小さいほど、処理優先度が高いものとする。また、図67において、丸付き数字は、対応するオブジェクトのpriorityの値を表している。

[0435] 図67の例では、オブジェクト#1乃至オブジェクト#4のpriorityが、それぞれ、1,2,3,4である場合、オブジェクト#1とオブジェクト#2が選択される。また、オブジェクト#1乃至オブジェクト#4のpriorityが、それぞれ、3,2,1,4に変更されると、オブジェクト#2とオブジェクト#3が選択される。さらに、オブジェクト#1乃至オブジェクト#4のpriorityが、それぞれ、3,4,1,2に変更されると、オブジェクト#3とオブジェクト#4が選択される。

[0436] 以上のように、表示領域内のオブジェクトのうちの、処理優先度の高いオブジェクトのobject audioのオーディオストリームのみを選択的に取得することで、Webサーバ142（212）と動画再生端末144（214）間の帯域を効率的に利用できる。contentkindに基づいてオブジェクトを選択する場合も同様である。

[0437] <第5実施の形態>
(トラックの概要)

図 6 8 は、本開示を適用した第 5 実施の形態におけるトラックの概要を説明する図である。

[0438] 図 6 8 に示すように、第 5 実施の形態では、各トラックが異なるファイル (3da_base.mp4/3da_channel.mp4/3da_object_1.mp4/3da_hoa.mp4/3da_meta.mp4) として記録される点が、第 2 実施の形態と異なっている。

[0439] 第 5 実施の形態における情報処理システムおよび情報処理システムによる処理は、第 4 実施の形態と同様であるので、説明は省略する。

[0440] <第 6 実施の形態>

(トラックの概要)

図 6 9 は、本開示を適用した第 6 実施の形態におけるトラックの概要を説明する図である。

[0441] 図 6 9 に示すように、第 6 実施の形態では、各トラックが異なるファイル (3da_basemeta.mp4/3da_channel.mp4/3da_object_1.mp4/3da_hoa.mp4) として記録される点が、第 3 実施の形態と異なっている。

[0442] 第 6 実施の形態における情報処理システムおよび情報処理システムによる処理は、オブジェクト位置情報を取得するために、Object metadata trackの代わりにBase trackのオーディオストリームが取得される点を除いて、第 4 実施の形態と同様であるので、説明は省略する。

[0443] なお、第 1 乃至第 3 実施の形態、第 5 実施の形態、および第 6 実施の形態においても、priorityやcontentkindに基づいて表示領域内のオブジェクトを選択することは可能である。

[0444] また、第 1 乃至第 6 実施の形態において、ストリーミング再生部は、図 23 のストリーミング再生部 120 のように、表示領域外のオブジェクトのオーディオストリームも取得し、そのオブジェクトのobject audioも合成して出力するようにしてもよい。

[0445] さらに、第 1 乃至第 6 実施の形態では、metadataからオブジェクト位置情報を取得したが、MPDファイルからオブジェクト位置情報を取得するようにしてもよい。

[0446] <3Dオーディオの階層構造の説明>

図70は、3Dオーディオの階層構造を示す図である。

[0447] 図70に示すように、3Dオーディオの音声データは、音声データごとに異なるオーディオエレメント (Element) とされる。オーディオエレメントのタイプとしては、SCE (Single Channel Element) とCPE (Channel Pair Element) がある。1チャンネル分の音声データのオーディオエレメントのタイプはSCEであり、2チャンネル分の音声データに対応するオーディオエレメントのタイプはCPEである。

[0448] オーディオエレメントは、同一の音声の種類 (Channel/Object/SAOC Objects/HOA) どうしでグループを形成する。グループタイプ (GroupType) としては、Channels, Objects, SAOC Objects, HOAなどがある。2以上のグループは、必要に応じて、switch Groupやgroup Presetを形成することができる。

[0449] switch Groupは、排他的に再生されるグループを定義するものである。即ち、図70に示すように、英語 (EN) 用のObject audioのグループと、フランス語 (FR) 用のObject audioのグループが存在する場合、いずれか一方のグループのみが再生されるべきである。従って、グループIDが2である英語用のObject audioのグループと、グループIDが3であるフランス語用のObject audioのグループとから、switch Groupが形成される。これにより、英語用のObject audioとフランス語用のObject audioが排他的に再生される。

[0450] 一方、group Presetは、コンテンツ製作者が意図するグループの組み合わせを定義するものである。

[0451] また、3Dオーディオのメタデータは、メタデータごとに異なるExtエレメント (Ext Element) とされる。Extエレメントのタイプとしては、Object Metadata, SAOC 3D Metadata, HOA Metadata, DRC Metadata, SpatialFrame, SaocFrameなどがある。Object MetadataのExtエレメントは、全てのObject audioのメタデータであり、SAOC 3D MetadataのExtエレメントは、全てのSAOC audioのメタデータである。また、HOA MetadataのExtエレメントは、全てのHOA audi

oのメタデータであり、DRC(Dinamic Range Control) MetadataのExtエレメントは、Object audio, SAOC audio、およびHOA audioの全てのメタデータである。

[0452] 以上のように、3Dオーディオのうちの音声データの分割単位としては、オーディオエレメント、グループタイプ、グループ、switch Group、およびgroup Presetがある。従って、音声データは、第1乃至第6実施の形態のようにグループタイプ（但し、object audioについてはオブジェクトごと）ごとにトラックに分割されるのではなく、オーディオエレメント、グループ、switch Group、またはgroup Presetごとに分割されるようにしてもよい。

[0453] また、3Dオーディオのうちのメタデータの分割単位としては、Extエレメントのタイプ（ExtElementType）、または、そのメタデータに対応するオーディオエレメントがある。従って、メタデータは、第1乃至第6実施の形態のようにExtエレメントのタイプごとに分割されるのではなく、そのメタデータに対応するオーディオエレメントごとに分割されるようにしてもよい。

[0454] なお、以下の説明では、音声データがオーディオエレメントごとに分割されるとともに、メタデータがExtエレメントのタイプごとに分割され、それぞれが異なるトラックのデータとして配置されるものとするが、他の単位で分割される場合も同様である。

[0455] <Webサーバの処理の第1の例の説明>

図71は、Webサーバ142（212）の処理の第1の例を説明する図である。

[0456] 図71の例では、ファイル生成装置141（211）からアップロードされる音声ファイルに対応する3Dオーディオが、5つのチャンネルのchannel audio、3つのオブジェクトのobject audio、および、そのobject audioのメタデータ（Object Metadata）からなる。

[0457] そして、5つのチャンネルのchannel audioが、前方中央（FC）のチャンネルのchannel audio、前方左右（FL,FR）のチャンネルのchannel audio、および後方左右（RL,RR）のチャンネルのchannel audioに分割され、それぞれ異

なるトラックのデータとして配置されている。また、各オブジェクトのobject audioが、それぞれ異なるトラックのデータとして配置されている。さらに、Object Metadataが、1つのトラックのデータとして配置されている。

[0458] また、図71に示すように、3Dオーディオの各オーディオストリームは、config情報とフレーム（サンプル）単位のデータからなる。そして、図71の例では、音声ファイルのオーディオストリームは、5つのチャンネルのchannel audio、3つのオブジェクトのobject audio、およびObject Metadataのconfig情報がまとめて配置されるとともに、各フレームのデータがまとめて配置される。

[0459] この場合、図71に示すように、Webサーバ142（212）は、ファイル生成装置141（211）からアップロードされた音声ファイルのオーディオストリームを、トラックごとに分割し、7つのトラックのオーディオストリームを生成する。具体的には、Webサーバ142（212）は、ssix boxなどの情報にしたがって、音声ファイルのオーディオストリームから、各トラックのconfig情報と音声データを抽出し、各トラックのオーディオストリームを生成する。各トラックのオーディオストリームは、そのトラックのconfig情報と各フレームの音声データからなる。

[0460] 図72は、Webサーバ142（212）のトラック分割処理を説明するフローチャートである。このトラック分割処理は、例えば、ファイル生成装置141（211）から音声ファイルがアップロードされたとき、開始される。

[0461] 図72のステップS441において、Webサーバ142（212）は、ファイル生成装置141からアップロードされた音声ファイルを格納する。

[0462] ステップS442において、Webサーバ142（212）は、音声ファイルのssix boxなどの情報にしたがって、音声ファイルを構成するオーディオストリームをトラックごとに分割する。

[0463] ステップS443において、Webサーバ142（212）は、各トラックのオーディオストリームを保持し、処理を終了する。このオーディオストリームは、動画再生端末144（214）の音声ファイル取得部192（264

）から要求されたとき、Webサーバ142（212）から動画再生端末144（214）に送信される。

[0464] <音声復号処理部の処理の第1の例の説明>

図73は、Webサーバ142（212）の処理が図71および図72で説明した処理である場合の音声復号処理部194の処理の第1の例を説明する図である。

[0465] 図73の例では、Webサーバ142（212）は、図71に示した各トラックのオーディオストリームを保持している。また、再生対象のトラックが、前方左右のチャンネルのchannel audio、後方左右のチャンネルのchannel audio、1つ目のオブジェクトのobject audio、およびObject Metadataのトラックである。これらのことは、後述する図75においても同様である。

[0466] この場合、音声ファイル取得部192（264）は、前方左右のチャンネルのchannel audio、後方左右のチャンネルのchannel audio、1つ目のオブジェクトのobject audio、およびObject Metadataのトラックのオーディオストリームを取得する。

[0467] 音声復号処理部194は、まず、音声ファイル取得部192（264）により取得されたObject Metadataのトラックのオーディオストリームから、1つ目のオブジェクトのobject audioのメタデータのオーディオストリームを抽出する。

[0468] 次に、図73に示すように、音声復号処理部194は、再生対象の音声のトラックのオーディオストリームと抽出されたメタデータのオーディオストリームとを合成する。具体的には、音声復号処理部194は、全てのオーディオストリームに含まれているConfig情報がまとめて配置されるとともに、各フレームのデータがまとめて配置されたオーディオストリームを生成する。そして、音声復号処理部194は、生成されたオーディオストリームを復号する。

[0469] 以上のように、再生対象のオーディオストリームが、1つのchannel audioのトラックのオーディオストリームのみではない場合、2以上のトラックの

オーディオストリームが再生対象となるため、復号前にオーディオストリームが合成される。

[0470] 一方、再生対象のオーディオストリームが、1つのchannel audioのトラックのオーディオストリームのみである場合、オーディオストリームを合成する必要はない。従って、音声復号処理部194は、音声ファイル取得部192（264）により取得されたオーディオストリームをそのまま復号する。

[0471] 図74は、Webサーバ142（212）の処理が図71および図72で説明した処理である場合の音声復号処理部194の復号処理の第1の例の詳細を説明するフローチャートである。この復号処理は、再生対象のトラックが1つのchannel audioのトラックのみではない場合に行われる図48のステップS229および図50のS287の処理の少なくとも一方の処理である。

[0472] 図74のステップS461において、音声復号処理部194は、生成するオーディオストリームに含まれるエレメントの数を表す全エレメント数に0を設定する。ステップS462において、音声復号処理部194は、生成するオーディオストリームに含まれるエレメントのタイプを表す全エレメント種別情報を初期化（クリア）する。

[0473] ステップS463において、音声復号処理部194は、再生対象のトラックのうちのまだ処理対象のトラックとされていないトラックを、処理対象のトラックとする。ステップS464において、音声復号処理部194は、処理対象のトラックのオーディオストリーム等から、処理対象のトラックに含まれるエレメントの数とタイプを取得する。

[0474] ステップS465において、音声復号処理部194は、取得されたエレメントの数を全エレメント数に加算する。ステップS466において、音声復号処理部194は、取得されたエレメントのタイプを、全エレメント種別情報に追加する。

[0475] ステップS467において、音声復号処理部194は、再生対象の全てのトラックを処理対象のトラックとしたかどうかを判定する。ステップS467でまだ再生対象の全てのトラックを処理対象のトラックとしていないと判

定された場合、処理はステップS 4 6 3に戻り、再生対象の全てのトラックを処理対象のトラックとするまで、ステップS 4 6 3乃至S 4 6 7の処理が繰り返される。

[0476] 一方、ステップS 4 6 7で再生対象の全てのトラックを処理対象のトラックとしたと判定された場合、処理はステップS 4 6 8に進む。ステップS 4 6 8において、音声復号処理部1 9 4は、全エレメント数と全エレメント種別情報を、生成するオーディオストリーム上の所定の位置に配置する。

[0477] ステップS 4 6 9において、音声復号処理部1 9 4は、再生対象のトラックのうちのまだ処理対象のトラックとされていないトラックを、処理対象のトラックとする。ステップS 4 7 0において、音声復号処理部1 9 4は、処理対象のトラックに含まれるエレメントのうちのまだ処理対象のエレメントとされていないエレメントを、処理対象のエレメントとする。

[0478] ステップS 4 7 1において、音声復号処理部1 9 4は、処理対象のトラックのオーディオストリームから処理対象のエレメントのConfig情報を取得し、生成するオーディオストリーム上に配置する。このとき、再生対象の全てのトラックの全てのエレメントのConfig情報は連続するように配置される。

[0479] ステップS 4 7 2において、音声復号処理部1 9 4は、処理対象のトラックに含まれる全てのエレメントを処理対象のエレメントとしたかどうかを判定する。ステップS 4 7 2でまだ全てのエレメントを処理対象のエレメントとしていないと判定された場合、処理はステップS 4 7 0に戻り、全てのエレメントを処理対象のエレメントとするまで、ステップS 4 7 0乃至S 4 7 2の処理が繰り返される。

[0480] 一方、ステップS 4 7 2で全てのエレメントを処理対象のエレメントとしたと判定された場合、処理はステップS 4 7 3に進む。ステップS 4 7 3において、音声復号処理部1 9 4は、再生対象の全てのトラックを処理対象のトラックとしたかどうかを判定する。ステップS 4 7 3でまだ再生対象の全てのトラックを処理対象のトラックとしていないと判定された場合、処理はステップS 4 6 9に戻り、再生対象の全てのトラックを処理対象のトラック

とするまで、ステップS 4 6 9乃至S 4 7 3の処理が繰り返される。

[0481] 一方、ステップS 4 7 3で再生対象の全てのトラックを処理対象のトラックとしたと判定された場合、処理はステップS 4 7 4に進む。ステップS 4 7 4において、音声復号処理部1 9 4は、処理対象のフレームを決定する。最初のステップS 4 7 4の処理では、先頭のフレームが処理対象のフレームに決定され、2回目以降のステップS 4 7 4の処理では、現在の処理対象のフレームの次のフレームが、新たな処理対象のフレームに決定される。

[0482] ステップS 4 7 5において、音声復号処理部1 9 4は、再生対象のトラックのうちのまだ処理対象のトラックとされていないトラックを、処理対象のトラックとする。ステップS 4 7 6において、音声復号処理部1 9 4は、処理対象のトラックに含まれるエレメントのうちのまだ処理対象のエレメントとされていないエレメントを、処理対象のエレメントとする。

[0483] ステップS 4 7 7において、音声復号処理部1 9 4は、処理対象のエレメントがEXTエレメントであるかどうかを判定する。ステップS 4 7 7で処理対象のエレメントがEXTエレメントではないと判定された場合、処理はステップS 4 7 8に進む。

[0484] ステップS 4 7 8において、音声復号処理部1 9 4は、処理対象のトラックのオーディオストリームから、処理対象のエレメントの処理対象のフレームの音声データを取得し、生成するオーディオストリーム上に配置する。このとき、再生対象の全てのトラックの全てのエレメントの同一のフレームのデータは連続するように配置される。ステップS 4 7 8の処理後、処理はステップS 4 8 1に進む。

[0485] 一方、ステップS 4 7 7で処理対象のエレメントがEXTエレメントであると判定された場合、処理はステップS 4 7 9に進む。ステップS 4 7 9において、音声復号処理部1 9 4は、処理対象のトラックのオーディオストリームから、処理対象のエレメントの処理対象のフレームの全てのオブジェクトのメタデータを取得する。

[0486] ステップS 4 8 0において、音声復号処理部1 9 4は、取得された全ての

オブジェクトのメタデータのうちの再生対象のオブジェクトのメタデータを、生成するオーディオストリーム上に配置する。このとき、再生対象の全てのトラックの全てのエレメントの同一のフレームのデータは連続するように配置される。ステップS 4 8 0の処理後、処理はステップS 4 8 1に進む。

[0487] ステップS 4 8 1において、音声復号処理部1 9 4は、処理対象のトラックに含まれる全てのエレメントを処理対象のエレメントとしたかどうかを判定する。ステップS 4 8 1でまだ全てのエレメントを処理対象のエレメントとしていないと判定された場合、処理はステップS 4 7 6に戻り、全てのエレメントを処理対象のエレメントとするまで、ステップS 4 7 6乃至S 4 8 1の処理が繰り返される。

[0488] 一方、ステップS 4 8 1で全てのエレメントを処理対象のエレメントとしたと判定された場合、処理はステップS 4 8 2に進む。ステップS 4 8 2において、音声復号処理部1 9 4は、再生対象の全てのトラックを処理対象のトラックとしたかどうかを判定する。ステップS 4 8 2でまだ再生対象の全てのトラックを処理対象のトラックとしていないと判定された場合、処理はステップS 4 7 5に戻り、再生対象の全てのトラックを処理対象のトラックとするまで、ステップS 4 7 5乃至S 4 8 2の処理が繰り返される。

[0489] 一方、ステップS 4 8 2で再生対象の全てのトラックを処理対象のトラックとしたと判定された場合、処理はステップS 4 8 3に進む。

[0490] ステップS 4 8 3において、音声復号処理部1 9 4は、全てのフレームを処理対象のフレームとしたかどうかを判定する。ステップS 4 8 3でまだ全てのフレームを処理対象のフレームとしていないと判定された場合、処理はステップS 4 7 4に戻り、全てのフレームを処理対象のフレームとするまで、ステップS 4 7 4乃至S 4 8 3の処理が繰り返される。

[0491] 一方、ステップS 4 8 3で全てのフレームを処理対象のフレームとしたと判定された場合、処理はステップS 4 8 4に進む。ステップS 4 8 4において、音声復号処理部1 9 4は、生成されたオーディオストリームを復号する。即ち、音声復号処理部1 9 4は、全エレメント数、全エレメント種別情報

、Config情報、および、音声データや再生対象のオブジェクトのメタデータが配置されたオーディオストリームを復号する。音声復号処理部194は、復号の結果得られる音声データ (Object audio, Channel audio, HOA audio) を音声合成処理部195に供給し、処理を終了する。

[0492] <音声復号処理部の処理の第2の例の説明>

図75は、Webサーバ142 (212) の処理が図71および図72で説明した処理である場合の音声復号処理部194の処理の第2の例を説明する図である。

[0493] 図75に示すように、音声復号処理部194の処理の第2の例は、生成するオーディオストリーム上に、全てのトラックのオーディオストリームを配置し、再生対象ではないトラックのオーディオストリームとしては復号結果がゼロになるストリームまたはフラグ (以下、ゼロストリームという) を配置する点が、第1の例と異なる。

[0494] 具体的には、音声ファイル取得部192 (264) は、Webサーバ142 (212) に保持されている全てのトラックのオーディオストリームに含まれるConfig情報と、再生対象のトラックのオーディオストリームに含まれる各フレームのデータを取得する。

[0495] 音声復号処理部194は、図75に示すように、全てのトラックのConfig情報を、生成するオーディオストリーム上にまとめて配置する。また、音声復号処理部194は、再生対象のトラックの各フレームのデータと、再生対象ではないトラックの各フレームのデータとしてのゼロストリームとを、生成するオーディオストリーム上にまとめて配置する。

[0496] 以上のように、音声復号処理部194は、生成するオーディオストリームに、再生対象ではないトラックのオーディオストリームとしてゼロストリームを配置するので、再生対象ではないオブジェクトのオーディオストリームも存在する。従って、生成するオーディオストリーム内に再生対象ではないオブジェクトのメタデータを含ませることが可能になる。よって、音声復号処理部194は、Object Metadataのトラックのオーディオストリームから、

再生対象のオブジェクトのメタデータのオーディオストリームを抽出する必要がなくなる。

[0497] なお、再生対象ではないトラックのConfig情報としては、ゼロストリームが配置されるようにしてもよい。

[0498] 図76は、Webサーバ142（212）の処理が図71および図72で説明した処理である場合の音声復号処理部194の復号処理の第2の例の詳細を説明するフローチャートである。この復号処理は、再生対象のトラックが1つのchannel audioのトラックのみではない場合に行われる図48のステップS229および図50のS287の処理の少なくとも一方の処理である。

[0499] 図76のステップS501およびS502の処理は、図74のステップS461およびS462の処理と同様であるので、説明は省略する。

[0500] ステップS503において、音声復号処理部194は、Webサーバ142（212）に保持されているオーディオストリームに対応するトラックのうちの、まだ処理対象のトラックとされていないトラックを、処理対象のトラックとする。

[0501] ステップS504乃至ステップS506の処理は、ステップS464乃至S466の処理と同様であるので、説明は省略する。

[0502] ステップS507において、音声復号処理部194は、Webサーバ142（212）に保持されているオーディオストリームに対応する全てのトラックを、処理対象のトラックとしたかどうかを判定する。ステップS507でまだ全てのトラックを処理対象のトラックとしていないと判定された場合、処理はステップS503に戻り、全てのトラックを処理対象のトラックとするまで、ステップS503乃至S507の処理が繰り返される。

[0503] 一方、ステップS507で全てのトラックを処理対象のトラックとしたと判定された場合、処理はステップS508に進む。ステップS508において、音声復号処理部194は、全エレメント数と全エレメント種別情報を、生成するオーディオストリーム上の所定の位置に配置する。

[0504] ステップS509において、音声復号処理部194は、Webサーバ142（

212) に保持されているオーディオストリームに対応するトラックのうちの、まだ処理対象のトラックとされていないトラックを、処理対象のトラックとする。ステップS510において、音声復号処理部194は、処理対象のトラックに含まれるエレメントのうちのまだ処理対象のエレメントとされていないエレメントを、処理対象のエレメントとする。

[0505] ステップS511において、音声復号処理部194は、処理対象のトラックのオーディオストリームから処理対象のエレメントのConfig情報を取得し、生成するオーディオストリーム上に配置する。このとき、Webサーバ142(212) に保持されているオーディオストリームに対応する全てのトラックの全てのエレメントのConfig情報は連続するように配置される。

[0506] ステップS512において、音声復号処理部194は、処理対象のトラックに含まれる全てのエレメントを処理対象のエレメントとしたかどうかを判定する。ステップS512でまだ全てのエレメントを処理対象のエレメントとしていないと判定された場合、処理はステップS510に戻り、全てのエレメントを処理対象のエレメントとするまで、ステップS510乃至S512の処理が繰り返される。

[0507] 一方、ステップS512で全てのエレメントを処理対象のエレメントとしたと判定された場合、処理はステップS513に進む。ステップS513において、音声復号処理部194は、Webサーバ142(212) に保持されているオーディオストリームに対応する全てのトラックを処理対象のトラックとしたかどうかを判定する。ステップS513でまだ全てのトラックを処理対象のトラックとしていないと判定された場合、処理はステップS509に戻り、全てのトラックを処理対象のトラックとするまで、ステップS509乃至S513の処理が繰り返される。

[0508] 一方、ステップS513で全てのトラックを処理対象のトラックとしたと判定された場合、処理はステップS514に進む。ステップS514において、音声復号処理部194は、処理対象のフレームを決定する。最初のステップS514の処理では、先頭のフレームが処理対象のフレームに決定され

、2回目以降のステップS 5 1 4の処理では、現在の処理対象のフレームの次のフレームが、新たな処理対象のフレームに決定される。

[0509] ステップS 5 1 5において、音声復号処理部1 9 4は、Webサーバ1 4 2（2 1 2）に保持されているオーディオストリームに対応するトラックのうちの、まだ処理対象のトラックとされていないトラックを、処理対象のトラックとする。

[0510] ステップS 5 1 6において、音声復号処理部1 9 4は、処理対象のトラックが再生対象のトラックであるかどうかを判定する。ステップS 5 1 6で処理対象のトラックが再生対象のトラックであると判定された場合、ステップS 5 1 7に進む。

[0511] ステップS 5 1 7において、音声復号処理部1 9 4は、処理対象のトラックに含まれるエレメントのうちのまだ処理対象のエレメントとされていないエレメントを、処理対象のエレメントとする。

[0512] ステップS 5 1 8において、音声復号処理部1 9 4は、処理対象のトラックのオーディオストリームから、処理対象のエレメントの処理対象のフレームの音声データを取得し、生成するオーディオストリーム上に配置する。このとき、Webサーバ1 4 2（2 1 2）に保持されているオーディオストリームに対応する全てのトラックの全てのエレメントの同一のフレームのデータは連続するように配置される。

[0513] ステップS 5 1 9において、音声復号処理部1 9 4は、処理対象のトラックに含まれる全てのエレメントを処理対象のエレメントとしたかどうかを判定する。ステップS 5 1 9でまだ全てのエレメントを処理対象のエレメントとしていないと判定された場合、処理はステップS 5 1 7に戻り、全てのエレメントを処理対象のエレメントとするまで、ステップS 5 1 7乃至S 5 1 9の処理が繰り返される。

[0514] 一方、ステップS 5 1 9で全てのエレメントを処理対象のエレメントとしたと判定された場合、処理はステップS 5 2 3に進む。

[0515] また、ステップS 5 1 6で処理対象のトラックが再生対象のトラックでは

ないと判定された場合、処理はステップS 5 2 0に進む。ステップS 5 2 0において、音声復号処理部1 9 4は、処理対象のトラックに含まれるエレメントのうちのまだ処理対象のエレメントとされていないエレメントを、処理対象のエレメントとする。

[0516] ステップS 5 2 1において、音声復号処理部1 9 4は、生成するオーディオストリーム上に、処理対象のエレメントの処理対象のフレームのデータとしてゼロストリームを、配置する。このとき、Webサーバ1 4 2 (2 1 2) に保持されているオーディオストリームに対応する全てのトラックの全てのエレメントの同一のフレームのデータは連続するように配置される。

[0517] ステップS 5 2 2において、音声復号処理部1 9 4は、処理対象のトラックに含まれる全てのエレメントを処理対象のエレメントとしたかどうかを判定する。ステップS 5 2 2でまだ全てのエレメントを処理対象のエレメントとしていないと判定された場合、処理はステップS 5 2 0に戻り、全てのエレメントを処理対象のエレメントとするまで、ステップS 5 2 0乃至S 5 2 2の処理が繰り返される。

[0518] 一方、ステップS 5 2 2で全てのエレメントを処理対象のエレメントとしたと判定された場合、処理はステップS 5 2 3に進む。

[0519] ステップS 5 2 3において、音声復号処理部1 9 4は、Webサーバ1 4 2 (2 1 2) に保持されているオーディオストリームに対応する全てのトラックを処理対象のトラックとしたかどうかを判定する。ステップS 5 2 2でまだ全てのトラックを処理対象のトラックとしていないと判定された場合、処理はステップS 5 1 5に戻り、再生対象の全てのトラックを処理対象のトラックとするまで、ステップS 5 1 5乃至S 5 2 3の処理が繰り返される。

[0520] 一方、ステップS 5 2 3で全てのトラックを処理対象のトラックとしたと判定された場合、処理はステップS 5 2 4に進む。

[0521] ステップS 5 2 4において、音声復号処理部1 9 4は、全てのフレームを処理対象のフレームとしたかどうかを判定する。ステップS 5 2 4でまだ全てのフレームを処理対象のフレームとしていないと判定された場合、処理は

ステップS 5 1 4に戻り、全てのフレームを処理対象のフレームとするまで、ステップS 5 1 4乃至S 5 2 4の処理が繰り返される。

- [0522] 一方、ステップS 5 2 4で全てのフレームを処理対象のフレームとしたと判定された場合、処理はステップS 5 2 5に進む。ステップS 5 2 5において、音声復号処理部1 9 4は、生成されたオーディオストリームを復号する。即ち、音声復号処理部1 9 4は、全エレメント数、全エレメント種別情報、並びにWebサーバ1 4 2 (2 1 2)に保持されているオーディオストリームに対応する全てのトラックのConfig情報およびデータが配置されたオーディオストリームを復号する。音声復号処理部1 9 4は、復号の結果得られる音声データ (Object audio, Channel audio, HOA audio) を音声合成処理部1 9 5に供給し、処理を終了する。

- [0523] <Webサーバの処理の第2の例の説明>

図7 7は、Webサーバ1 4 2 (2 1 2)の処理の第2の例を説明する図である。

- [0524] 図7 7のWebサーバ1 4 2 (2 1 2)の処理の第2の例は、各オブジェクトのObject Metadataが、それぞれ異なるトラックのデータとして音声ファイルに配置される点を除いて、図7 1の第1の例と同一である。

- [0525] 従って、図7 7に示すように、Webサーバ1 4 2 (2 1 2)は、ファイル生成装置1 4 1 (2 1 1)からアップロードされた音声ファイルのオーディオストリームを、トラックごとに分割し、9つのトラックのオーディオストリームを生成する。

- [0526] この場合のWebサーバ1 4 2 (2 1 2)のトラック分割処理は、図7 2のトラック分割処理と同様であるので、説明は省略する。

- [0527] <音声復号処理部の処理の第3の例の説明>

図7 8は、Webサーバ1 4 2 (2 1 2)の処理が図7 7で説明した処理である場合の音声復号処理部1 9 4の処理を説明する図である。

- [0528] 図7 8の例では、Webサーバ1 4 2 (2 1 2)は、図7 7に示した各トラックのオーディオストリームを保持している。また、再生対象のトラックが、

前方左右のチャンネルのchannel audio、後方左右のチャンネルのchannel audio、1つ目のオブジェクトのobject audio、および1つ目のオブジェクトのObject Metadataのトラックである。

[0529] この場合、音声ファイル取得部192(264)は、前方左右のチャンネルのchannel audio、後方左右のチャンネルのchannel audio、1つ目のオブジェクトのobject audio、および1つ目のオブジェクトのObject Metadataのトラックのオーディオストリームを取得する。音声復号処理部194は、取得された再生対象のトラックのオーディオストリームを合成し、生成されたオーディオストリームを復号する。

[0530] 以上のように、Object Metadataが、オブジェクトごとに異なるトラックのデータとして配置される場合、音声復号処理部194は、再生対象のオブジェクトのObject Metadataのオーディオストリームを抽出する必要がない。従って、音声復号処理部194は、復号するオーディオストリームを容易に生成することができる。

[0531] 図79は、Webサーバ142(212)の処理が図77で説明した処理である場合の音声復号処理部194の復号処理の詳細を説明するフローチャートである。この復号処理は、再生対象のトラックが1つのchannel audioのトラックのみではない場合に行われる図48のステップS229および図50のS287の処理の少なくとも一方の処理である。

[0532] 図79の復号処理は、ステップS477、S479およびS480の処理が行われない点、および、ステップS478の処理で音声データだけでなくメタデータも配置される点を除いて、図74の復号処理と同様である。即ち、図79のステップS541乃至S556の処理は、図74のステップS461乃至S476と同様であり、図79のステップS557の処理では、ステップS478の処理と同様に、処理対象の要素の処理対象のフレームのデータが配置される。また、ステップS558乃至S561の処理は、図74のステップS481乃至S484の処理と同様である。

[0533] なお、上述した説明では、動画再生端末144(214)が復号するオー

ディオストリームを生成したが、Webサーバ142（212）が、再生対象のトラックの組み合わせとして想定される組み合わせのオーディオストリームを生成するようにしてもよい。この場合、動画再生端末144（214）は、再生対象のトラックの組み合わせのオーディオストリームをWebサーバ142（212）から取得し、復号するだけで、再生対象のトラックの音声を再生することができる。

[0534] また、音声復号処理部194は、Webサーバ142（212）から取得された再生対象のトラックのオーディオストリームを、トラックごとに復号するようにしてもよい。この場合、音声復号処理部194は、復号の結果得られる音声データとメタデータを合成する必要がある。

[0535] <ベースサンプルのシンタクスの第2の例>

（ベースサンプルに配置されるConfig情報のシンタクスの第2の例）

図80は、ベースサンプルに配置されるConfig情報のシンタクスの第2の例を示す図である。

[0536] 図80の例では、Config情報として、ベースサンプルに配置されるエレメントの数（numElements）が記述される。また、ベースサンプルに配置される各エレメントのタイプ(usacElementType)として、Extエレメントを表す「ID_USAC_EXT」が記述されるとともに、各エレメントのExtエレメント用のConfig情報（mpegh3daExtElementConfig）が記述される。

[0537] 図81は、図80のExtエレメント用のConfig情報（mpegh3daExtElementConfig）のシンタクスの例を示す図である。

[0538] 図81に示すように、図80のExtエレメント用のConfig情報（mpegh3daExtElementConfig）としては、ExtエレメントのタイプとしてExtractorを表す「ID_EXT_ELE_EXTRACTOR」が記述される。また、Extractor用のConfig情報（ExtractorConfig）が記述される。

[0539] 図82は、図81のExtractor用のConfig情報（ExtractorConfig）のシンタクスの例を示す図である。

[0540] 図82に示すように、図81のExtractor用のConfig情報（ExtractorConf

g) としては、そのExtractorが参照するエレメントのタイプ（種別）（usac Element Type Extractor）が記述される。また、エレメントのタイプ（usac Element Type Extractor）がExtエレメントを表す「ID_USAC_EXT」である場合、そのExtエレメントのタイプ(usacExtElementTypeExtractor)が記述される。さらに、参照するエレメント（サブサンプル）のConfig情報のサイズ（size）（configLength）と位置（configOffset）が記述される。

[0541] （ベースサンプルに配置されるフレーム単位のデータのシンタクスの第2の例）

図83は、ベースサンプルに配置されるフレーム単位のデータのシンタクスの第2の例を示す図である。

[0542] 図83に示すように、ベースサンプルに配置されるフレーム単位のデータとしては、そのデータのエレメントであるExtエレメントのタイプとしてExtractorを表す「ID_EXT_ELE_EXTRACTOR」が記述される。また、Extractorのデータ（Extractor Metadata）が記述される。

[0543] 図84は、図83のExtractorのデータ（Extractor Metadata）のシンタクスの例を示す図である。

[0544] 図84に示すように、図83のExtractorのデータ（Extractor Metadata）としては、そのExtractorが参照するエレメントのデータのサイズ（elementLength）と位置（elementOffset）が記述される。

[0545] <ベースサンプルのシンタクスの第3の例>

（ベースサンプルに配置されるConfig情報のシンタクスの第3の例）

図85は、ベースサンプルに配置されるConfig情報のシンタクスの第3の例を示す図である。

[0546] 図85の例では、Config情報として、ベースサンプルに配置されるエレメントの数（numElements）が記述される。また、Config情報を配置するサンプルがExtractorであるかどうかを示すExtractorフラグ（flag Extractor）として、Extractorであることを示す1が記述される。また、elementLengthPresentとして1が記述される。

[0547] また、ベースサンプルに配置される各エレメントのタイプ(usacElementType)として、そのエレメントが参照するエレメントのタイプが記述される。エレメントのタイプ(usacElementType)がExtエレメントを表す「ID_USAC_EXT」である場合、Extエレメントのタイプ(usacExtElementType)が記述される。さらに、参照するエレメントのConfig情報のサイズ (size) (configLength) と位置 (configOffset) が記述される。

[0548] (ベースサンプルに配置されるフレーム単位のデータのシンタクスの第3の例)

図86は、ベースサンプルに配置されるフレーム単位のデータのシンタクスの第3の例を示す図である。

[0549] 図86に示すように、ベースサンプルに配置されるフレーム単位のデータとしては、そのデータが参照するエレメントのデータのサイズ (elementLength) と位置 (elementOffset) が記述される。

[0550] <第7実施の形態>

(オーディオストリームの構成例)

図87は、本開示を適用した情報処理システムの第7実施の形態において音声ファイルに格納されるオーディオストリームの構成例を示す図である。

[0551] 図87に示すように、第7実施の形態では、音声ファイルには、各グループタイプ(但し、object audioについてはオブジェクトごと)の3Dオーディオのサンプル単位の符号化データが、サブサンプルとして配置されるオーディオストリーム(3D audio stream)が格納される。

[0552] また、音声ファイルには、各グループタイプの3Dオーディオのサンプル単位の符号化データのサイズ、位置、およびグループタイプを含むextractorを、サブサンプルとするヒントストリーム(3D audio hint stream)が格納される。extractorの構成は、上述した構成と同様であり、グループタイプがextractorの種別として記述される。

[0553] (トラックの概要)

図88は、第7実施の形態におけるトラックの概要を説明する図である。

[0554] 図 8 8 に示すように、第 7 実施の形態では、オーディオストリームとヒントストリームに異なるトラックが割り当てられる。オーディオストリームのトラックのTrack Referenceとしては、対応するヒントストリームのトラックのトラックID「2」が記述される。また、ヒントストリームのトラックのTrack Referenceとしては、対応するオーディオストリームのトラックのトラックID「1」が記述される。

[0555] オーディオストリームのトラックのサンプルエントリのシンタクスは、図 3 4 に示したシンタクスであり、ヒントストリームのトラックのサンプルエントリのシンタクスは、図 3 5 乃至図 3 8 に示したシンタクスを含む。

[0556] (ファイル生成装置の処理の説明)

図 8 9 は、第 7 実施の形態におけるファイル生成装置のファイル生成処理を説明するフローチャートである。

[0557] なお、第 7 実施の形態におけるファイル生成装置は、音声符号化処理部 1 7 1 と音声ファイル生成部 1 7 2 の処理が異なる点を除いて、図 4 5 のファイル生成装置 1 4 1 と同一である。従って、以下では、第 7 実施の形態におけるファイル生成装置、音声符号化処理部、音声ファイル生成部を、ファイル生成装置 3 0 1、音声符号化処理部 3 4 1、音声ファイル生成部 3 4 2 という。

[0558] 図 8 9 のステップ S 6 0 1 乃至 S 6 0 5 の処理は、図 4 6 のステップ S 1 9 1 乃至 S 1 9 5 の処理と同様であるので、説明は省略する。

[0559] ステップ S 6 0 6 において、音声符号化処理部 3 4 1 は、外部から入力される動画コンテンツの 3 D オーディオをグループタイプごとに符号化し、図 8 7 のオーディオストリームを生成する。音声符号化処理部 3 4 1 は、生成されたオーディオストリームを音声ファイル生成部 3 4 2 に供給する。

[0560] ステップ S 6 0 7 において、音声ファイル生成部 3 4 2 は、音声符号化処理部 3 4 1 から供給されるオーディオストリームから、サブサンプル情報を取得する。サブサンプル情報とは、各グループタイプの 3 D オーディオのサンプル単位の符号化データのサイズ、位置、およびグループタイプである。

[0561] ステップS 6 0 8において、音声ファイル生成部3 4 2は、サブサンプル情報に基づいて、図8 7のヒントストリームを生成する。ステップS 6 0 9において、音声ファイル生成部3 4 2は、オーディオストリームとヒントストリームを異なるトラックとして多重化し、音声ファイルを生成する。このとき、音声ファイル生成部3 4 2は、外部から入力される画枠サイズ情報をサンプルエントリに格納する。音声ファイル生成部3 4 2は、生成された音声ファイルをMPD生成部1 7 3に供給する。

[0562] ステップS 6 1 0およびS 6 1 1の処理は、図4 6のステップS 1 9 9およびS 2 0 0の処理と同様であるので、説明は省略する。

[0563] (動画再生端末の処理の説明)

図9 0は、第7実施の形態における動画再生端末のストリーミング再生部のオーディオ再生処理を説明するフローチャートである。

[0564] なお、第7実施の形態におけるストリーミング再生部は、MPD処理部1 9 1、音声ファイル取得部1 9 2、および音声復号処理部1 9 4の処理が異なる点、並びに、音声選択部1 9 3が設けられない点を除いて、図4 7のストリーミング再生部1 9 0と同一である。従って、以下では、第7実施の形態におけるストリーミング再生部、MPD処理部、音声ファイル取得部、および音声復号処理部を、ストリーミング再生部3 6 0、MPD処理部3 8 1、音声ファイル取得部3 8 2、および音声復号処理部3 8 3という。

[0565] 図9 0のステップS 6 2 1において、ストリーミング再生部3 6 0のMPD処理部3 8 1は、MPD取得部9 1から供給されるMPDファイルを解析し、再生対象のセグメントの音声ファイルのURL等の情報を取得して、音声ファイル取得部3 8 2に供給する。

[0566] ステップS 6 2 2において、音声ファイル取得部3 8 2は、MPD処理部3 8 1から供給されるURL等の情報に基づいて、再生対象のセグメントのInitial SegmentをWebサーバに要求し、取得する。

[0567] ステップS 6 2 3において、音声ファイル取得部3 8 2は、Initial Segment内のmoovボックスのヒントストリームのトラック（以下では、ヒントトラ

ックという)のサンプルエントリから、参照トラックであるオーディオストリームのトラックのトラックIDを取得する。

- [0568] ステップS 6 2 4において、音声ファイル取得部3 8 2は、MPD処理部3 8 1から供給されるURL等の情報に基づいて、再生対象のセグメントのmedia segmentの先頭からsidxボックスとssixボックスをWebサーバに要求し、取得する。
- [0569] ステップS 6 2 5において、音声ファイル取得部3 8 2は、ステップS 6 2 4で取得されたsidxボックスとssixボックスから、ヒントトラックの位置情報を取得する。
- [0570] ステップS 6 2 6において、音声ファイル取得部3 8 2は、ステップS 6 2 5で取得されたヒントトラックの位置情報に基づいて、ヒントストリームをWebサーバに要求し、取得する。そして、音声ファイル取得部3 8 2は、ヒントストリームから、再生対象の3 Dオーディオのグループタイプのextractorを取得する。なお、再生対象の3 Dオーディオがobject audioである場合、再生対象とするオブジェクトは、画枠サイズ情報とオブジェクト位置情報とに基づいて選択される。
- [0571] ステップS 6 2 7において、音声ファイル取得部3 8 2は、ステップS 6 2 4で取得されたsidxボックスとssixボックスから、参照トラックの位置情報を取得する。ステップS 6 2 8において、音声ファイル取得部3 8 2は、ステップS 6 2 7で取得された参照トラックの位置情報と、取得されたextractorに含まれるサブサンプル情報とに基づいて、再生対象の3 Dオーディオのグループタイプのオーディオストリームの位置情報を決定する。
- [0572] ステップS 6 2 9において、音声ファイル取得部3 8 2は、ステップS 6 2 7で決定された位置情報に基づいて、再生対象の3 DオーディオのグループタイプのオーディオストリームをWebサーバに要求し、取得する。音声ファイル取得部3 8 2は、取得されたオーディオストリームを音声復号処理部3 8 3に供給する。
- [0573] ステップS 6 3 0において、音声復号処理部3 8 3は、音声ファイル取得

部 3 8 2 から供給されるオーディオストリームを復号し、その結果得られる音声データを音声合成処理部 1 9 5 に供給する。

[0574] ステップ S 6 3 1 において、音声合成処理部 1 9 5 は、音声データを出力し、処理を終了する。

[0575] なお、第 7 実施の形態では、オーディオストリームのトラックとヒントトラックが同一の音声ファイルに格納されるようにしたが、異なるファイルに格納されるようにしてもよい。

[0576] <第 8 実施の形態>

(トラックの概要)

図 9 1 は、本開示を適用した情報処理システムの第 8 実施の形態におけるトラックの概要を説明する図である。

[0577] 第 8 実施の形態の音声ファイルは、格納されるヒントストリームがグループタイプごとのストリームである点が、第 7 実施の形態の音声ファイルと異なる。即ち、第 8 実施の形態のヒントストリームは、グループタイプごとに生成され、各ヒントストリームには、各グループタイプの 3 D オーディオのサンプル単位の符号化データのサイズ、位置、およびグループタイプを含む extractor が、サンプルとして配置される。なお、3 D オーディオに複数のオブジェクトの object audio が含まれる場合、extractor は、オブジェクトごとにサブサンプルとして配置される。

[0578] また、図 9 1 に示すように、第 8 実施の形態では、オーディオストリームと各ヒントストリームに異なるトラックが割り当てられる。オーディオストリームのトラックは、図 8 8 のオーディオストリームのトラックと同一であるため、説明は省略する。

[0579] グループタイプ「Channels」、「Objects」、「HOA」、「metadata」のヒントトラックの Track Reference としては、対応するオーディオストリームのトラックのトラック ID 「1」が記述される。

[0580] グループタイプ「Channels」、「Objects」、「HOA」、および「metadata」のヒントトラックのサンプルエントリのシンタクスは、それぞれ、サンプ

ルエントリの種別を表す情報を除いて、図 3 5 乃至図 3 8 に示したシンタクスと同一である。グループタイプ「Channels」、「Objects」、「HOA」、および「metadata」のヒントトラックのサンプルエントリの種別を表す情報は、図 3 5 乃至図 3 8 の情報の数字の 1 が 2 に代わったものであり、この 2 は、ヒントトラックのサンプルエントリであることを表す。

[0581] (音声ファイルの構成例)

図 9 2 は、音声ファイルの構成例を示す図である。

[0582] 図 9 2 に示すように、音声ファイルには、図 9 1 に示した全てのトラックが格納される。即ち、音声ファイルには、オーディオストリームと各グループタイプのヒントストリームが格納される。

[0583] 第 8 実施の形態におけるファイル生成装置のファイル生成処理は、図 8 7 のヒントストリームの代わりに、グループタイプごとのヒントストリームが生成される点を除いて、図 8 9 のファイル生成処理と同様である。

[0584] また、第 8 実施の形態における動画再生端末のストリーミング再生部のオーディオ再生処理は、ステップ S 6 2 3 において、参照トラックのトラック ID だけでなく、再生対象のグループタイプのヒントトラックのトラック ID を取得する点、ステップ S 6 2 5 において再生対象のグループタイプのヒントトラックの位置情報を取得する点、およびステップ S 6 2 6 において再生対象のグループタイプのヒントストリームを取得する点を除いて、図 9 0 のオーディオ再生処理と同様である。

[0585] なお、第 8 実施の形態では、オーディオストリームのトラックとヒントトラックが同一の音声ファイルに格納されるようにしたが、異なるファイルに格納されるようにしてもよい。

[0586] 例えば、図 9 3 に示すように、オーディオストリームのトラックが 1 つの音声ファイル (3D audio stream MP4 File) に格納され、ヒントトラックが 1 つの音声ファイル (3D audio hint stream MP4 File) に格納されるようにしてもよい。また、図 9 4 に示すように、ヒントトラックは、複数の音声ファイルに分割して格納されるようにしてもよい。図 9 4 の例では、各ヒント

トラックが異なる音声ファイルに格納されている。

- [0587] また、第8実施の形態では、グループタイプがobjectsである場合であってもグループタイプごとにヒントストリームが生成されたが、グループタイプがobjectsである場合、オブジェクトごとにヒントストリームが生成されるようにしてもよい。この場合、各オブジェクトのヒントストリームに対して異なるトラックが割り当てられる。
- [0588] 以上のように、第7および第8実施の形態の音声ファイルでは、全ての3Dオーディオのオーディオストリームが1つのトラックとして格納される。従って、動画再生端末は、そのトラックを取得することにより、全ての3Dオーディオのオーディオストリームを再生することができる。
- [0589] また、第7および第8実施の形態の音声ファイルでは、ヒントストリームが格納される。従って、動画再生端末は、サブサンプルと、サブサンプルのサイズや位置とを対応付けたテーブルが記述されるmoofボックスを参照せずに、全ての3Dオーディオのオーディオストリームのうちの、所望のグループタイプのオーディオストリームのみを取得し、再生することができる。
- [0590] さらに、第7および第8実施の形態の音声ファイルでは、全ての3Dオーディオのオーディオストリームとヒントストリームを格納するだけで、動画再生端末にグループタイプごとにオーディオストリームを取得させることができる。従って、グループタイプごとのオーディオストリームの取得を可能にするために、放送やローカルストレージの用途として生成された全ての3Dオーディオのオーディオストリームとは別に、グループタイプごとの3Dオーディオのオーディオストリームを用意する必要がない。
- [0591] なお、第7および第8実施の形態では、グループタイプごとにextractorが生成されたが、extractorの生成単位は、オーディオエレメント、グループ、switch Group、およびgroup Presetであってもよい。
- [0592] extractorの生成単位がグループである場合、第8実施の形態の各ヒントトラックのサンプルエントリには、対応するグループに関する情報が含まれる。グループに関する情報は、グループのID、グループに分類されるエレメ

ントのデータの内容を表す情報等により構成される。グループがswitch Groupを形成する場合、そのグループのヒントトラックのサンプルエントリには、そのswitch Groupに関する情報も含まれる。switch Groupに関する情報は、switch GroupのID、switch Groupを形成するグループのID等により構成される。第7実施の形態のヒントトラックのサンプルエントリには、第8実施の形態の全てのヒントトラックのサンプルエントリに含まれる情報が含まれる。

[0593] また、第7および第8実施の形態におけるセグメント構造は、図39および図40のセグメント構造と同一である。

[0594] <第9実施の形態>

(本開示を適用したコンピュータの説明)

上述したWebサーバの一連の処理は、ハードウェアにより実行することもできるし、ソフトウェアにより実行することもできる。一連の処理をソフトウェアにより実行する場合には、そのソフトウェアを構成するプログラムが、コンピュータにインストールされる。ここで、コンピュータには、専用のハードウェアに組み込まれているコンピュータや、各種のプログラムをインストールすることで、各種の機能を実行することが可能な、例えば汎用のパーソナルコンピュータなどが含まれる。

[0595] 図95は、上述したWebサーバの一連の処理をプログラムにより実行するコンピュータのハードウェアの構成例を示すブロック図である。

[0596] コンピュータにおいて、CPU (Central Processing Unit) 601, ROM (Read Only Memory) 602, RAM (Random Access Memory) 603は、バス604により相互に接続されている。

[0597] バス604には、さらに、入出力インタフェース605が接続されている。入出力インタフェース605には、入力部606、出力部607、記憶部608、通信部609、及びドライブ610が接続されている。

[0598] 入力部606は、キーボード、マウス、マイクロフォンなどよりなる。出力部607は、ディスプレイ、スピーカなどよりなる。記憶部608は、ハ

ードディスクや不揮発性のメモリなどよりなる。通信部609は、ネットワークインタフェースなどよりなる。ドライブ610は、磁気ディスク、光ディスク、光磁気ディスク、又は半導体メモリなどのリムーバブルメディア611を駆動する。

[0599] 以上のように構成されるコンピュータでは、CPU601が、例えば、記憶部608に記憶されているプログラムを、入出力インタフェース605及びバス604を介して、RAM603にロードして実行することにより、上述した一連の処理が行われる。

[0600] コンピュータ（CPU601）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611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。

[0601] コンピュータでは、プログラムは、リムーバブルメディア611をドライブ610に装着することにより、入出力インタフェース605を介して、記憶部608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、通信部609で受信し、記憶部608にインストールすることができる。その他、プログラムは、ROM602や記憶部608に、あらかじめインストールしておくことができる。

[0602] なお、コンピュータ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良いし、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良い。

[0603] また、上述した動画再生端末のハードウェア構成は、図95のコンピュータと同様の構成にすることができる。この場合、例えば、CPU601が、制御用ソフトウェア161（221）、動画再生ソフトウェア162（222）、およびアクセス用ソフトウェア163（223）を実行する。動画再生端末144（214）の処理は、ハードウェアにより実行することもできる。

[0604] 本明細書において、システムとは、複数の構成要素（装置、モジュール（部品）等）の集合を意味し、すべての構成要素が同一筐体中にあるか否かは問わない。したがって、別個の筐体に収納され、ネットワークを介して接続されている複数の装置、及び、1つの筐体の中に複数のモジュールが収納されている1つの装置は、いずれも、システムである。

[0605] なお、本開示の実施の形態は、上述した実施の形態に限定されるものではなく、本開示の要旨を逸脱しない範囲において種々の変更が可能である。

[0606] 例えば、ファイル生成装置141（211）は、タイル単位で画像ファイルを生成するのではなく、全てのタイルの符号化データを多重化してビデオストリームを生成し、1つの画像ファイルを生成するようにしてもよい。

[0607] 本開示は、MPEG-H 3D audioだけでなく、Object毎にstreamを作成することのできるaudio codec全般に適応可能である。

[0608] また、本開示は、ストリーミング再生ではなく、放送やローカルストレージ再生を行う情報処理システムにも適用することができる。

[0609] さらに、本開示は、以下のような構成もとることができる。

[0610] （1）

複数の種類の音声データが前記種類に応じて複数のトラックに分割されて配置されるファイルのうちの所定の前記トラックの音声データを取得する取得部

を備える情報処理装置。

（2）

前記種類は、前記音声データのエレメント、前記エレメントのタイプ、または前記エレメントが分類されるグループである

ように構成された

前記（1）に記載の情報処理装置。

（3）

前記取得部により取得された前記所定のトラックの前記音声データを復号する復号部

をさらに備える

前記（１）または（２）に記載の情報処理装置。

（４）

前記復号部は、前記所定のトラックの数が複数である場合、前記取得部により取得された前記所定のトラックの前記音声データを合成し、合成後の前記音声データを復号する

前記（３）に記載の情報処理装置。

（５）

前記ファイルには、複数のオブジェクト単位の音声データが前記オブジェクトごとに異なる前記トラックに分割されて配置されるとともに、全ての前記オブジェクト単位の音声データのメタデータが前記トラックとは異なるトラックにまとめて配置され、

前記取得部は、再生対象のオブジェクトの前記トラックの前記音声データを、前記所定のトラックの前記音声データとして取得するとともに、前記メタデータを取得し、

前記復号部は、前記取得部により取得された前記メタデータから、前記再生対象のオブジェクトの前記メタデータを抽出し、そのメタデータと前記取得部により取得された前記音声データとを合成する

ように構成された

前記（４）に記載の情報処理装置。

（６）

前記ファイルには、複数のオブジェクト単位の音声データが前記オブジェクトごとに異なる前記トラックに分割されて配置されるとともに、全ての前記オブジェクト単位の音声データのメタデータが前記トラックとは異なるトラックにまとめて配置され、

前記取得部は、再生対象のオブジェクトの前記トラックの前記音声データを、前記所定のトラックの前記音声データとして取得するとともに、前記メタデータを取得し、

前記復号部は、再生対象ではない前記トラックの前記音声データとしての復号結果がゼロになるゼロデータと、前記取得部により取得された前記音声データおよび前記メタデータとを合成する

ように構成された

前記（４）に記載の情報処理装置。

（７）

前記ファイルには、複数のオブジェクト単位の音声データが前記オブジェクトごとに異なる前記トラックに分割されて配置されるとともに、前記オブジェクト単位の音声データのメタデータが前記オブジェクトごとに異なるトラックに配置され、

前記取得部は、再生対象のオブジェクトの前記トラックの前記音声データを、前記所定のトラックの前記音声データとして取得するとともに、再生対象の前記オブジェクトのメタデータを取得し、

前記復号部は、前記取得部により取得された前記音声データと前記メタデータを合成する

ように構成された

前記（４）に記載の情報処理装置。

（８）

前記複数のトラックの前記音声データは、１つの前記ファイルに配置される

ように構成された

前記（１）乃至（７）のいずれかに記載の情報処理装置。

（９）

前記複数のトラックの前記音声データは、前記トラックごとに異なる前記ファイルに配置される

ように構成された

前記（１）乃至（７）に記載の情報処理装置。

（１０）

前記ファイルには、前記複数の種類の前記音声データに関する情報が、前記複数のトラックとは異なるトラックとして配置される

ように構成された

前記（１）乃至（９）のいずれかに記載の情報処理装置。

（１１）

前記複数の種類の前記音声データに関する情報は、前記音声データに対応する画像データの画枠サイズを示す画枠サイズ情報を含む

ように構成された

前記（１０）に記載の情報処理装置。

（１２）

前記ファイルには、前記複数のトラックとは異なるトラックの前記音声データとして、その音声データに対応する他の前記トラックの前記音声データの位置を示す情報が配置される

ように構成された

前記（１）乃至（９）のいずれかに記載の情報処理装置。

（１３）

前記ファイルには、前記複数のトラックとは異なるトラックの前記データとして、そのデータに対応する他の前記トラックの前記音声データの位置を示す情報と、他の前記トラックの前記音声データのメタデータが配置される

ように構成された

前記（１）乃至（９）のいずれかに記載の情報処理装置。

（１４）

前記音声データのメタデータは、前記音声データの取得位置を示す情報を含む

ように構成された

前記（１３）に記載の情報処理装置。

（１５）

前記ファイルは、前記トラックの他の前記トラックとの参照関係を表す情

報を含む

ように構成された

前記（１）乃至（１４）のいずれかに記載の情報処理装置。

（１６）

前記ファイルは、各トラックの前記音声データのコーデック情報を含む

ように構成された

前記（１）乃至（１５）のいずれかに記載の情報処理装置。

（１７）

所定の前記種類の音声データは、他の前記種類の音声データの取得位置を示す情報である

ように構成された

前記（１）乃至（１６）のいずれかに記載の情報処理装置。

（１８）

情報処理装置が、

複数の種類の音声データが前記種類に応じて複数のトラックに分割されて配置されるファイルのうちの所定の前記トラックの音声データを取得する取得ステップ

を含む情報処理方法。

（１９）

複数の種類の音声データが前記種類に応じて複数のトラックに分割されて配置されるファイルを生成する生成部

を備える情報処理装置。

（２０）

情報処理装置が、

複数の種類の音声データが前記種類に応じて複数のトラックに分割されて配置されるファイルを生成する生成ステップ

を含む情報処理方法。

符号の説明

[0611] 141 ファイル生成装置, 144 動画再生端末, 172 音声ファイル生成部, 192 音声ファイル取得部, 193 音声選択部, 211 ファイル生成装置, 214 動画再生端末, 241 音声ファイル生成部, 264 音声ファイル取得部

請求の範囲

- [請求項1] 複数の種類の音声データが前記種類に応じて複数のトラックに分割されて配置されるファイルのうちの所定の前記トラックの音声データを取得する取得部
を備える情報処理装置。
- [請求項2] 前記種類は、前記音声データの要素、前記要素のタイプ、または前記要素が分類されるグループである
ように構成された
請求項1に記載の情報処理装置。
- [請求項3] 前記取得部により取得された前記所定のトラックの前記音声データを復号する復号部
をさらに備える
請求項1に記載の情報処理装置。
- [請求項4] 前記復号部は、前記所定のトラックの数が複数である場合、前記取得部により取得された前記所定のトラックの前記音声データを合成し、合成後の前記音声データを復号する
請求項3に記載の情報処理装置。
- [請求項5] 前記ファイルには、複数のオブジェクト単位の音声データが前記オブジェクトごとに異なる前記トラックに分割されて配置されるとともに、全ての前記オブジェクト単位の音声データのメタデータが前記トラックとは異なるトラックにまとめて配置され、
前記取得部は、再生対象のオブジェクトの前記トラックの前記音声データを、前記所定のトラックの前記音声データとして取得するとともに、前記メタデータを取得し、
前記復号部は、前記取得部により取得された前記メタデータから、前記再生対象のオブジェクトの前記メタデータを抽出し、そのメタデータと前記取得部により取得された前記音声データとを合成する
ように構成された

請求項 4 に記載の情報処理装置。

[請求項6]

前記ファイルには、複数のオブジェクト単位の音声データが前記オブジェクトごとに異なる前記トラックに分割されて配置されるとともに、全ての前記オブジェクト単位の音声データのメタデータが前記トラックとは異なるトラックにまとめて配置され、

前記取得部は、再生対象のオブジェクトの前記トラックの前記音声データを、前記所定のトラックの前記音声データとして取得するとともに、前記メタデータを取得し、

前記復号部は、再生対象ではない前記トラックの前記音声データとしての復号結果がゼロになるゼロデータと、前記取得部により取得された前記音声データおよび前記メタデータとを合成する

ように構成された

請求項 4 に記載の情報処理装置。

[請求項7]

前記ファイルには、複数のオブジェクト単位の音声データが前記オブジェクトごとに異なる前記トラックに分割されて配置されるとともに、前記オブジェクト単位の音声データのメタデータが前記オブジェクトごとに異なるトラックに配置され、

前記取得部は、再生対象のオブジェクトの前記トラックの前記音声データを、前記所定のトラックの前記音声データとして取得するとともに、再生対象の前記オブジェクトのメタデータを取得し、

前記復号部は、前記取得部により取得された前記音声データと前記メタデータを合成する

ように構成された

請求項 4 に記載の情報処理装置。

[請求項8]

前記複数のトラックの前記音声データは、1つの前記ファイルに配置される

ように構成された

請求項 1 に記載の情報処理装置。

- [請求項9] 前記複数のトラックの前記音声データは、前記トラックごとに異なる前記ファイルに配置される
ように構成された
請求項1に記載の情報処理装置。
- [請求項10] 前記ファイルには、前記複数の種類の前記音声データに関する情報が、前記複数のトラックとは異なるトラックとして配置される
ように構成された
請求項1に記載の情報処理装置。
- [請求項11] 前記複数の種類の前記音声データに関する情報は、前記音声データに対応する画像データの画枠サイズを示す画枠サイズ情報を含む
ように構成された
請求項10に記載の情報処理装置。
- [請求項12] 前記ファイルには、前記複数のトラックとは異なるトラックの前記音声データとして、その音声データに対応する他の前記トラックの前記音声データの位置を示す情報が配置される
ように構成された
請求項1に記載の情報処理装置。
- [請求項13] 前記ファイルには、前記複数のトラックとは異なるトラックの前記データとして、そのデータに対応する他の前記トラックの前記音声データの位置を示す情報と、他の前記トラックの前記音声データのメタデータが配置される
ように構成された
請求項1に記載の情報処理装置。
- [請求項14] 前記音声データのメタデータは、前記音声データの取得位置を示す情報を含む
ように構成された
請求項13に記載の情報処理装置。
- [請求項15] 前記ファイルは、前記トラックの他の前記トラックとの参照関係を

表す情報を含む

ように構成された

請求項 1 に記載の情報処理装置。

[請求項16] 前記ファイルは、各トラックの前記音声データのコーデック情報を含む

ように構成された

請求項 1 に記載の情報処理装置。

[請求項17] 所定の前記種類の音声データは、他の前記種類の音声データの取得位置を示す情報である

ように構成された

請求項 1 に記載の情報処理装置。

[請求項18] 情報処理装置が、

複数の種類の音声データが前記種類に応じて複数のトラックに分割されて配置されるファイルのうちの所定の前記トラックの音声データを取得する取得ステップ

を含む情報処理方法。

[請求項19] 複数の種類の音声データが前記種類に応じて複数のトラックに分割されて配置されるファイルを生成する生成部

を備える情報処理装置。

[請求項20] 情報処理装置が、

複数の種類の音声データが前記種類に応じて複数のトラックに分割されて配置されるファイルを生成する生成ステップ

を含む情報処理方法。